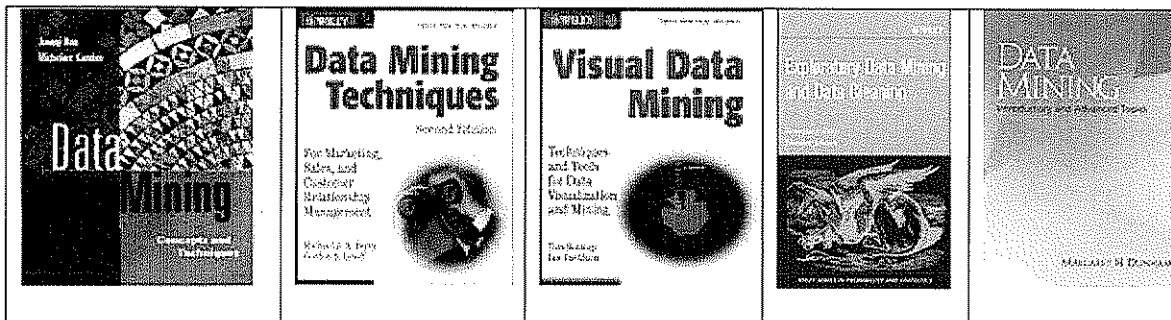




MIST812 – Decision Support Systems **STAT820 – Decision Support Systems**

First Semester 2005

Unit Outline



Student Resource Centre
Division of Economic & Financial Studies
MACQUARIE UNIVERSITY

Students in this unit should read this unit outline carefully at the start of semester. It contains important information about the unit. If anything in it is unclear, please consult one of the teaching staff in the unit.

ABOUT THIS UNIT

Decision support systems provide information to support semi-structured and unstructured decisions in modern organisations. The decision support systems studied in this unit are computer based and provide for user manipulation of source data extracted from databases both internal and external to the organisation. Students will study decision support systems involving: data warehousing and data marts, online analytic processing, data mining and the geographic information systems. The unit will be taught using modern software including SPSS Clementine for data mining and MapInfo for the GIS package.

Mist812/Stat820 is a four credit point unit offered by the Statistics department in the Division of Economic and Financial Studies. It can be taken either day in the first semester or evening in the second semester. This unit expands on topics covered in MIST800 *Computer Applications in Business and available to graduate students both in Business and in Applied Statistics. It is offered as part of the Master of Commerce/Master of International Business Program.*

Software:

SPSS, SPSS-CLEMENTINE, MAPINFO, Microsoft Excel & Access

Prerequisite:

Basic quantitative knowledge (STAT170 or MIST800 or equivalent).

TEACHING STAFF

Lecturer In Charge: Associate Professor Julian Leslie

Room: C5C484

Phone: 9850 85931

e-mail jleslie@efs.mq.edu.au

GIS

Professor Don McNeil

Room: C5C 473

Phone: 9850 8553

e-mail: dmcneil@efs.mq.edu.au

Data Mining

Dr Ayse Bilgin

Room: C5C458

Phone: 9850 8509

e-mail: abilgin@efs.mq.edu.au

Tutors:

Ms. Ling Li

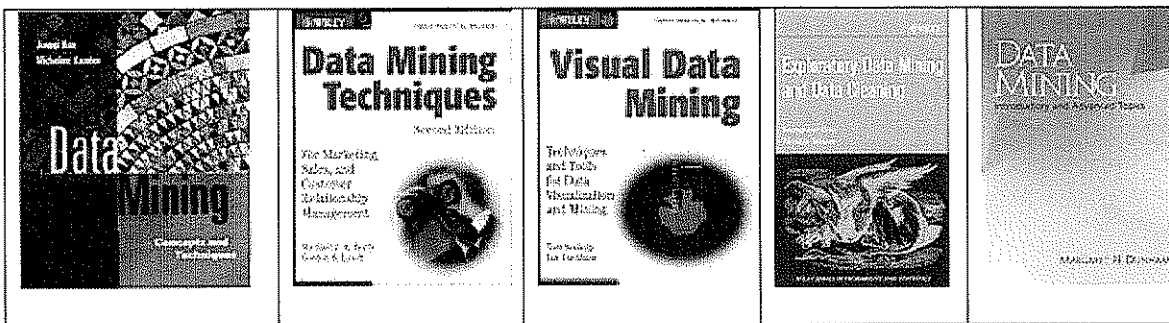
Ms. Sri Dewi

RECOMMENDED TEXT BOOK

Data Mining: Concepts and techniques by Jiawei Han and Micheline Kamber, 2001, Morgan and Kaufmann (library call number QA76.9.D343.H36 2001)

RECOMMENDED REFERENCE TEXTS

- 1) Data mining techniques for marketing, sales and customer relationship management by Michael Berry and Gordon Linoff, 2004, John Wiley (library call number HF5415.125 .B47 2004)
- 2) Visual Data Mining: Techniques and Tools for Data Visualization and Mining by Tom Soukup, Ian Davidson, May 2002, Wiley and Sons (it is available in the library)
- 3) Exploratory Data Mining and Data Cleaning by Tamraparni Dasu, Theodore Johnson, May 2003 (library call number QA76.9.D343 D34 2003)
- 4) Data mining introductory and advanced topics by Margaret Dunham, 2003, Prentice Hall/Pearson Education (library call number: QA76.9.D343.D86 2003)



CLASSES

Lectures

Lectures begin in Week 1. Students should attend **ONE** 2-hour session per week: Mondays between 2-pm and 4pm in C5C209.

Tutorials

Tutorials also begin in Week 1. The aim of tutorials is to practise techniques learnt in lectures. They are designed so that students work through the exercises and ask as many questions as they need to improve their understanding. Tutors are the facilitators in the tutorial groups. They will assist students and instead of giving them straight answers for every question, they will create an environment for thinking process and discussion between the students.

The timetable for classes can be found on the University web site at:
<http://www.timetables.mq.edu.au/>

UNIT WEB PAGE

Information relating to this unit can be found by visiting the Macquarie University Statistics Department web site. The URL for this unit is
<http://www.stat.mq.edu.au/units/stat820/>

LEARNING OUTCOMES

By the end of this unit students should be able to:

- have an understanding of the principles and the concepts of the data mining
- understand and create concept hierarchies
- use market basket analysis to improve the sales of a given company
- use classification and cluster analysis as data mining tools
- understand how the decision trees are developed and be able to interpret the output of the decision trees
- organise data which is suitable to display as a map
- create active earth maps with hyperlinks
- use MapInfo Software to create region boundaries for a chosen region
- create contour maps and Voronoi polygons
- to understand the link between data mining, geographical information systems and good decision making

GENERIC SKILLS

University study aims, not only to provide you with knowledge and skills in a particular academic discipline, but also to equip you with some generic skills. By the end of this unit students should:

- have improved their ability to work co-operatively as a team member
- have enhanced their problem solving ability
- have improved their written communication skills, particularly report writing skills
- have enhanced their critical thinking skills
- be confident in the use of different software packages for solving problems

TEACHING AND LEARNING STRATEGY

- students are expected to attend all the lectures and the tutorials
- in data mining section, weekly practical exercises are set for individual assessment of lab tasks. Therefore, if students decide to work in a group, the final outcome should be written individually
- the GIS group project will be prepared as a group and presented by every member of the group
- if for any reason, students can not hand in their assessment tasks on time, they have to contact one of the teaching staff in advance
- students should hand in and collect their marked papers from ERIC (Economic Resource & Information Centre) C5C244
- the solutions to lab exercises will not be given out however individual help for each student will be given during the consultation hours

ASSESSMENT

Seven weekly lab exercises related to data mining are due at BEGINNING of your lecture session on week following date of issue (e.g. Week 2 lab exercise solution is due in Week 3 before the lecture or by 2pm). You need to hand them into the appropriate box in ERIC (C5C240). These weekly lab exercises will be corrected by your tutor. There will not be any marks for these weekly exercises. It is important to collect them from ERIC so that you can make the corrections and prepare a better assignment than otherwise possible.

Seven weekly lab assignments are due at the beginning of your lecture session two weeks after they are issued. You need to hand them into the appropriate box in ERIC (C5C240). These assignments will be marked. This seven weekly lab exercises will be included into the final grading where each one of them will be scaled to 3% (**21%** of the final grade all together). You can also collect them from ERIC.

After the first two weeks, you will be put into groups for the GIS project. If you want to change your group, you have to find a student from the group that you want to move into who would swap with you. If you can not find a student to swap places, you have to remain in the group that you have been allocated to. If you could find a student who is willing to swap with you please inform Don McNeil or Ayse Bilgin.

Group Assignment report (**10%**) related to GIS is due in week 8 (2 pm Monday May 2). Presentation time table will be given to you on week 7.

Mid semester test (14%) will be held in the first 30 minutes of the week 7 lecture. This test covers the first six weeks of lecture material and readings. Students may bring one A4 sized sheet of notes, formulas, etc., which may be written on both sides and is easily readable (at least 8pt sized font). This summary must be submitted with your test paper. The rest of the lecture will cover new work.

Final examination (55%) is 2 hours 30 minutes long with 10 minutes reading time. This exam will be held on Monday 20 June 2005, between 2.00 pm – 4:40pm in Price Theatre, and will examine any material covered throughout the course. The examination is 'closed book'. You may refer only to a single self-prepared A4 sheet of crib notes which may be written on both sides and must be easily readable (at least 8pt sized font). This summary must be submitted with your exam paper and is marked for conforming to the guidelines given. Any other materials such as lecture notes and textbooks are not permitted.

Calculators are permitted, but may be used only as calculators, and not as storage devices. Mobile phones should be disabled.

Full marks may be obtained by correctly answering all the questions in the exam. Marks allocated to the questions will be given on the cover page of exam paper.

NOTE: To obtain a passing grade, both coursework and exam performance must be satisfactory.

OVERALL ASSESSMENT

Students are expected to gain a reasonable level of proficiency in weekly topics.

The overall assessment for MIST812 – STAT820 is:

Seven weekly lab assignments	21%
GIS Group Project	10%
Mid semester test	14%
Final examination	55%

The mark (SNG) recorded for this unit will be based on the weighted components above.

RELATIONSHIP BETWEEN ASSESSMENT AND LEARNING OUTCOMES

While attendance at classes is important it is only a small proportion of the total workload for the unit: reading, research in the library, working with other students in groups, completing assignments, using the computer and private study are all part of the work involved. At Macquarie it is expected that the average student should spend four hours per week per credit point.

You are expected to present yourself for examination at the time and place designated in the University Examination Timetable. Your final exam will be on week 15, 2pm in Price Theatre.

The only exception to not sitting an examination at the designated time is because of documented illness or unavoidable disruption. In these circumstances you may wish to consider applying for Special Consideration. Information about unavoidable disruption and the special consideration process is available at <http://www.reg.mq.edu.au/Forms/APSCon.pdf>

If a Supplementary Examination is granted as a result of the Special Consideration process the examination will be scheduled after the conclusion of the official examination period.

You are advised that it is Macquarie University policy not to set early examinations for individuals or groups of students. All students are expected to ensure that they are available until the end of the teaching semester that is the final day of the official examination period.

STUDENT SUPPORT SERVICES

Macquarie University provides a range of Academic Student Support Services. Details of these services can be accessed at <http://www.student.mq.edu.au>.

UNIVERSITY POLICY ON GRADING

Academic Senate has a set of guidelines on the distribution of grades across the range from fail to high distinction. Your final result will include one of these grades plus a standardised numerical grade (SNG).

On occasion your raw mark for a unit (i.e., the total of your marks for each assessment item) may not be the same as the SNG which you receive. Under the Senate guidelines, results may be scaled to ensure that there is a degree of comparability across the university, so that units with the same past performances of their students should achieve similar results.

It is important that you realise that the policy does not require that a minimum number of students are to be failed in any unit. In fact it does something like the opposite, in requiring examiners to explain their actions if more than 20% of students fail in a unit.

The process of scaling does not change the order of marks among students. A student who receives a higher raw mark than another will also receive a higher final scaled mark.

The grades and what they mean are given as below:

- HD – High Distinction** Denotes a performance that meets all unit objectives in such an exceptional way and with such marked excellence that it deserves the highest level of recognition.
- D – Distinction** Denotes performance that clearly deserves a very high level of recognition as an excellent achievement in the unit.
- C – Credit** Denotes performance that is substantially better than would normally be expected of competent students in the unit.
- P – Pass** Denotes performance that satisfies unit objectives.
- PC – Conceded Pass** Denotes performance that meets unit objectives only marginally.
- F – Fail** Denotes that a candidate has failed to complete a unit satisfactorily.

For further explanation of the policy see

<http://www.mq.edu.au/senate/MQUonly/Issues/Guidelines2003.doc> or
<http://www.mq.edu.au/senate/MQUonly/Issues/detailedguidelines.doc>.

PLAGIARISM

The University defines plagiarism in its rules: "Plagiarism involves using the work of another person and presenting it as one's own." Plagiarism is a serious breach of the University's rules and carries significant penalties. You must read the University's practices and procedures on plagiarism. These can be found in the *Handbook of Undergraduate Studies* or on the web at: <http://www.student.mq.edu.au/plagiarism/>

The policies and procedures explain what plagiarism is, how to avoid it, the procedures that will be taken in cases of suspected plagiarism, and the penalties if you are found guilty. Penalties may include a deduction of marks, failure in the unit, and/or referral to the University Discipline Committee.

ADVANCED STATISTICS COMPUTER LABS AND THEIR CONDITIONS OF USE

We have two labs that students can use during the term: W6B301 and Statistical Information Systems Lab (SIS Lab). Both of these labs are fitted with surveillance cameras and the activities in the room are recorded for security purposes. The first lab is located in building W6B and the SIS Lab is located in the Becton-Dickinson (BD) Building.

W6B301 OPENING HOURS

The lab is open 24 hours per day 7 days per week.

WARNING: students are strongly advised not to remain alone in the room after normal office hours.

You are encouraged to phone University Security

- **phone x7112 from inside the lab,**
- **see <http://www.bgo.mq.edu.au/security2.htm>)**

at any time after hours, during term time, if you require an escort to your vehicle or public transport.

WHILE USING W6B301 YOU MUST COMPLY WITH ANY REQUEST BY MACQUARIE UNIVERSITY SECURITY STAFF.

To gain access to W6B301 after hours

Each student must obtain a swipe-card which is to be used at entry points to the building W6B and (as of mid-year) for entry into the lab in room W6B301. But until mid-year there will be a press button door lock on Room 301 the code for which will be disclosed in lectures.

For the swipe-card, there is a form available from the Dept. Administrator Ms Sue Pe (Room C5C474) or it may be issued in lectures. This has to be completed and returned to your lecturer or to Ms Sue Pe. The forms are entered on a security database and then cards will be programmed for

each student. Students are to collect their cards from the front desk of the Division of Economic & Financial Studies. Ask for Ms Hiranya Loader or Ms Sue Coleman who will be able to issue the cards.

IF YOU LOSE YOUR CARD YOU WILL BE CHARGED \$50 FOR A REPLACEMENT.

NOTE:

- you must NOT lend your card to anyone
- you must carry your student ID with you when you use the card
- the campus security officers know the ID's of those students who are currently in the lab. They will periodically check the ID's of students in the lab. *Anyone without their student ID or who does not have his or her **own** swipe-card will be told to leave the building immediately*
- *swipe-cards that have been borrowed from another student will be immediately confiscated and no replacement card will be issued*

To gain access to the SIS Lab

The SIS Lab is in the BD building near the junction of Research Park Road and the road/path running down the centre of the University (past the Library, C5C, the Council Building, E7A, etc.). It is a 5 minute walk from C5C.

Check the map on the website:

http://www.bgo.mq.edu.au/images/S_E.gif

The SIS Lab is on the second floor. As you get out of the lift, turn right then right again and it is the door on the left. If it is locked during weekdays from 9 – 5, contact the dept. (9850 8555) or Julian Leslie (9850 8593). This lab has about 14 computers running Windows and 4 running both Unix and Windows.

NOTE: The lab is to be used for doing lab exercises, assignments and projects arising *only from units that use W6B301 for lectures/tutorials or prac classes.*

SIS LAB OPENING HOURS

The SIS Lab is open only during the hours of 9 am – 6 pm on weekdays. IT IS NOT AVAILABLE OUTSIDE THESE HOURS. STUDENTS MUST LEAVE THE LAB BY 6pm. UNDER NO CIRCUMSTANCES IS A COURSEWORK MASTERS STUDENT TO REMAIN IN THIS LAB AFTER HOURS.

Obtaining User Account in these labs

Each student will be given a user name and password for these labs once they are listed as enrolled in a MIST unit. After the first time logging into the server, the students need to change their password. The new (changed) password will expire in 30 days and needs to be changed again. If you do not change your password, you will not be able to login to the

server again. If this happens, please talk to your tutor or the computer lab administrator:

Mr. Alfred Wong, awong@efs.mq.edu.au

A time-table for the classes scheduled for each week will be displayed on the door of W6B301. If there is a class in progress, students who are not enrolled in that class are not allowed to use the computers in the lab without permission from the tutor.

PROBLEMS WITH LAB COMPUTERS?

Problems with lab computers (W6B or SISL) should be reported as follows:

1. if the problem occurs during a class report problem to your tutor
2. if problem occurs outside class time, then report problem by phone or e-mail to

Ms Susan Pe (ext 8550 sdjunaid@efs.mq.edu.au)
or by e-mail to the lab administrator

Mr Alfred Wong awong@efs.mq.edu.au

(USING YOUR MU E-MAIL BROWSER ACCOUNT and no other – staff are instructed to ignore e-mails from Hotmail accounts, etc). BE SURE TO INCLUDE YOUR NAME AND CLASS, THE LAB AND PC NUMBER AND A BRIEF DESCRIPTION OF THE PROBLEM.

WEEKLY READINGS

The weekly readings will be given to each student as part of the lecture notes and they are examinable.

FURTHER INTERESTING READINGS

Diego Kuonen (2003) "Is Data Mining for Gold `Statistical déjà vu'?" *Bulletin of the Swiss Statistical Society*, 44: pages 5-9.

Source http://www.statoo.com/en/publications/2003.DM.SSS.44/2003_DM_SSS_44.pdf

An Introduction to Data Mining

Source <http://www.thearling.com/text/dmwhite/dmwhite.htm>

Outliers in Statistical Data (Chapter 1 and chapter 13) by Vic Barnett and Toby Lewis, John Wiley and Sons, Third Edition, 1994:Library Call Number:Qa276.B2849/1994

Applications of Data Mining to Electronic Commerce

by RON KOHAVI and FOSTER PROVOST, *Data Mining and Knowledge Discovery*, 5, 5–10, 2001; Kluwer Academic Publishers. Manufactured in The Netherlands.

(Source

<http://citeseer.ist.psu.edu/cache/papers/cs/22262/http:zSzzSzrobotics.stanford.eduzSzerszSzronnykzSzecommerce-dmzSzeditorial.pdf/kohavi01applications.pdf>)

Data Mining for Fun and Profit, David J. Hand; Gordon Blunt; Mark G. Kelly; Niall M. Adams, *Statistical Science*, Vol.15, No.2 (May 2000), 111-126. (Search on <http://www.jstor.org/search/>)

Data Mining Components for Direct Marketing: The Tale of Two DMs by Sergei Ananyan (Source <http://www.megaputer.com/tech/wp/com.php3>)

Data Mining by Michael C. Lovell, *The Review of Economics and Statistics*, Vol.65, No.1 (Feb, 1983), 1-12.

(Source

http://econpapers.hhs.se/article/tprrestat/v_3A65_3Ay_3A1983_3Ai_3A1_3Ap_3A1-12.htm OR Search on <http://www.jstor.org/search/>)

Issues for On-line Analytical Mining of Data Warehouses, by Jiawei Han, Sonny H.S. Chee and Jenny Y. Chiang, *Proc. of 1998 SIGMOD'96 Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'98)*, Seattle, Washington, June 1998, pp. 2:1-2:5. (Source <ftp://ftp.fas.sfu.ca/pub/cs/han/pdf/dmkd98.pdf>)

Data mining on the web, by Dan R. Grenning

(Source <http://www.newarchitectmag.com/archives/2000/01/greening/>)

Principles of data mining, David J Hand, Hikki Mannila and Padhraic Smyth, August 2001, The MIT Press.

(Source <http://mitpress.mit.edu/catalog/item/default.asp?sid=B0DFD3D8-24E9-47DC-862B-2954CDCC66FF&ttype=2&tid=3520&mode=toc>)

Data mining cookbook : modeling data for marketing, risk and customer relationship management, by Olivia Parr Rud, Wiley and Sons, 2001, (Library call number: QA76.9.D343 R83 2001)

Clementine (SPSS) success stories (Source <http://www.spss.com/success/>)

IMPEDIMENTS TO EXPLORATORY DATA MINING SUCCESS

(Source <http://www.zsolutions.com/pdfs/Impediments.pdf>)

**MIST812/STAT820 Decision Support Systems
UNIT SCHEDULE**

WEEK	LECTURE TOPIC	Recommended reading	Assessment Due
28 Feb W1	Data Mining: Principles and Concepts	Chap1, Han & Kamber	
7 Mar W2	Data Mining: Data Warehouse and OLAP Technology	Chap2, Han & Kamber	Lab Ex 1
14 Mar W3	Data Mining: Data Preprocessing (Missing values, noisy data, inconsistent data) Data Mining: Concept hierarchies, interestingness measures & visualisation	Chap3 and Chap4 Han & Kamber	Lab Ex 2 Lab Ass 1
21 Mar W4	GIS: Introduction – presenting business geographic data		Lab Ex 3 Lab Ass 2
28 Mar W5	Public Holiday (no lectures or tutorials)		
4 Apr W6	GIS: Active earth maps		Lab Ass 3
11 Apr W7	GIS: Maps with hyperlinks		
SEMESTER BREAK: 16 April – 1 May			
2 May W8	GIS: Contour maps		GIS group project
9 May W9	Data Mining: Descriptive data mining	Chap5, Han & Kamber	
16 May W10	Data Mining: Mining Association Rules - Market Basket Analysis	Chap6, Han & Kamber	Lab Ex 9
23 May W11	Data Mining: Classification and Prediction	Chap7, Han & Kamber	Lab Ex 10 Lab Ass 9
30 May W12	Data Mining: Cluster Analysis	Chap8, Han & Kamber	Lab Ex 11 Lab Ass 10
6 June W13	Practice Exam Paper will be discussed during the lecture. There are no tutorials.		Lab Ex 12 Lab Ass 11 Lab Ass 12

Note that all DM lab exercises are due by 2pm in ERIC C5C244