COMP9318
Data Warehousing and Data Mining
Course Introduction

Course Details

- Course Code/Title : COMP9318 Data Warehousing and Data Mining
- Units of Credit : 6
- Course Website : http://www.cse.unsw.edu.au/~cs9318
- Handbook Entry : http://www.handbook.unsw.edu.au/postgraduate/courses/2011/COMP9318.html

Course Aims

This course aims to introduce the foundation of data warehousing, the theories of various data mining techniques and explore the practice of developing data mining applications.

Learning Outcomes

At the end of this course, students should be able to:

- familiar with major data mining tasks and paradigms.
- able to formulate a practical problem into a DM problem and solve it.
- understand the algorithmic challenges and pitfalls in solving DM problems.

Assessments

There will be 2 projects, 2 assignments, and final exam.

Assignments (50%):

- Ass1 - Week 2 - Week 6 (25%) : Data Warehouse + Data Cleaning
- Ass2 - Week 7 - Week 12 (25%) : Data Mining Techniques

How you should submit your work will be indicated each assignment.

Note: Penalty for later ass submission: **0** marks to any later submission

Projects (50%):

- Proj1 - Week 4 - 7 (25%)
- Proj2 - Week 8 - 11 (25%)

How you should submit your work will be indicated in each project.

Note: Penalty for later proj submission: 10% reduction for the 1st day, another 30% reduction for each following day.

Final Exam (100%):
There will be a 3-hour exam in the exam period.

Note: If you are ill on the day of the exam, **do not attend** the exam. We will **not** accept medical special consideration claims from people who have already attempted the exam.

Calculating Your Final Marks (Harmonic Mean)

```
Mark1 = ass1 + ass2 + proj1 + proj2
Mark2 = marks for final exam
Final Mark = 2 * Mark1 * Mark2 / (Mark1 + Mark2)
```

Course Administration

Course Staff

Name	Office	Phone	E-mail	Role
Xuemin Lin	K17-503	56493	lxue@cse	lecturer in charge
Zhitao Shen	K17-201	56861	shenz@cse	course admin

Note: You are invited to meet us **in person** during consultation time slots and during the lab periods. You are also welcome to contact us via e-mail if some thing is urgent.

Lectures Time

This course has a 3-hour lecture per week and helds on each Monday from 2pm - 5pm at Webster Theatre A (F Hall A)

(K-G14-190).

Course Resources : Textbooks

1. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 1st/2nd edition, Kaufmann Publishers.
2. Pang-Ning Tan, Michael Steinbach and Vipin Kumar *Introduction to Data Mining*, Addison-Wesley.

Course Resources : References

1. I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 2001.
2. R. Agrawal, J. Han, and H. Mannila, *Readings in Data Mining: A Database Perspective*, Morgan Kaufmann (in preparation).
3. U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996
4. U. Fayyad, G. Grinstein, and A. Wierse, *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann, 2001
5. J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2001
6. D. J. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*, MIT Press, 2001
7. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, 2001
8. T. M. Mitchell, *Machine Learning*, McGraw Hill, 1997
9. G. Piatetsky-Shapiro and W. J. Frawley. *Knowledge Discovery in Databases*. AAAI/MIT Press, 1991
10. S. M. Weiss and N. Indurkhya, *Predictive Data Mining, Morgan Kaufmann, 1998*

Academic Honesty and Plagiarism

Plagiarism *is defined as "using the words or ideas of others and presenting them as your own". UNSW and CSE treat plagiarism as academic misconduct, which means that it carries penalties as severe as being excluded from further study at UNSW. There are several on-line sources to help you understand what plagiarism is and how it is dealt with at UNSW:*

- *Learning Centre: Plagiarism and Academic Integrity*
- *MyUNSW: Plagiarism and Academic Misconduct*
- *CSE: Addendum to UNSW Plagiarism Guidelines*
- *CSE: Yellow Form (whose terms you have agreed to)*

Course Evaluation and Development

This course is evaluated each session using the CATEI system. In this session, we will use more concrete examples to demonstrate difficult concepts.

More Information

For further information on this course, and to keep up to date with any changes, please consult the course web site (frequently):

`http://www.cse.unsw.edu.au/~cs9318/`