UNIVERSITY OF NEBRASKA AT OMAHA
COURSE SYLLABUS/DESCRIPTION

Department and Course Number	ISQA/CSCI 4890
Course Title	Data Warehousing and Data Mining
Course Coordinator	Yong Shi
Total Credits	3
Date of Last Revision	6/26/15

1.0 Course Description:

1.1 Overview of content and purpose of the course (Catalog description).
This course provides students theoretical issues as well as practical methods for designing and constructing data warehouse and implementing data mining. After covering the essential concepts, issues, techniques to build an effective data warehouse, this course emphasizes the various techniques of data mining, such as association, classification, clustering and prediction for big data analyses. This course also promotes students to conduct a real-life big data analyzing project.

1.2 For whom course is intended.
This course is intended for undergraduate management information systems (MIS) and computer science (CS) majors who choose Data Mining and Business Intelligence concentration. It is also recommended to a senior undergraduate level in MIS or CS.

1.3 Prerequisites of the course (Courses)
- 1.3.1 ISQA 3310 or CSCI 4850

1.4 Prerequisites of the course (Topics).
- 1.4.1 The conceptual and practical foundation to database management systems
- 1.4.2 Relational Database Systems
- 1.4.3 The hierarchical models
- 1.4.4 The network models
- 1.4.5 Basic SQL

1.5 Unusual circumstances of the course.
None

2.0 Objectives:

2.1 Understand the fundamental principles of data warehousing and data mining
2.2 Describe the difference between transactional databases and data warehouses
2.3 Describe the relationship between OLAP and Data Warehousing
2.4 Use SQL to manage data in data warehousing and data mining
2.5 Build a data mart
2.6 Design and implement a data warehouse

2.7 Carry out data integration and transformation
2.8 Use and create simple data mining algorithms
2.9 Apply some commercial tools to conduct data mining
2.10 Interpret of data mining results

3.0 Content and Organization:

		Hours
3.1	Introduction to Data Warehousing and Data Mining	3
3.2	Data Warehouse and OLAP	3
3.3	Data Preprocessing	3
3.4	Data Mining Basis	3
3.5	Data Characterization and Comparison	4
3.6	Mining Association Rules	4
3.7	Classification and Prediction	6
3.8	Cluster Analysis	4
3.9	Integrated Methods of Data Mining	3
3.10	Big Data Analytics and Applications	3

4.0 Teaching Methodology:

4.1 Methods to be used.
The basic teaching method will be lectures, discussion, real-life cases, guest speakers. The existing software from the literature or produced by the instructor and commercial products will be used in the class.

4.2 Student role in the course
The students will attend lectures, participate in discussion in the class, complete assignments and projects, and complete all exams. The projects can be either individual or group work depending on the situation of each class.

4.3 Contact hours.
Three (3) hours per week.

5.0 Evaluation:

5.1 Type of student projects that will be the basis for evaluating student performance, specifying distinction between undergraduate and graduate, if applicable. For Laboratory projects, specify the number of weeks spent on each project).
The student products will be the assignments, group discussion, projects, examinations, and project presentations.

5.2 Basis for determining the final grade (Course requirements and grading standards) specifying distinction between undergraduate and graduate, if applicable.

Components	Grading
Participation	5%
Homework	10%
Projects	35%
Exams	50%

5.3 Grading scale and criteria.

Points	Grade
98-100	A+
95-97	A
90-94	A-
88-89	B+
85-87	B
80-84	B-
78-79	C+
75-77	C
70-74	C-
68-69	D+
65-67	D
60-64	D-
<=59	F

6.0 Resource Material

6.1 Textbooks and/or other required readings used in course.

The following textbooks, as examples, can be the required readings in the course. The instructor can also use the cited references in these books as supplementary materials for the course. There is no specific text for the course.

6.1.1 David Olson and Yong Shi, Introduction to Business Data Mining, McGraw-Hill, 2007.

6.1.2 J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2006.

6.2 Other suggested reading materials, if any.

6.2.1 Agrawal, R., Imielinski, T. and Swami, A. (1993) 'Data mining: A performance perspective', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5, 914-925.

6.2.2 Bajgier, S. M. and Hill, A. V. (1982) 'An experimental comparison of statistical and linear programming approaches to the discriminant problems', *Decision Sciences*, Vol.13, 604-618.

6.2.3 Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984), *Classification and Regression Trees*, Wadswroth, Belmont.

6.2.4 Carpenter, G.A., Grossberg, S., Reynolds, J.H., Markuzon, N., and Rosen, D.B. (1992) 'Fuzzy ARTMAP: A Neural Network Architecture for Incremental

Supervised Learning of Analog Multidimensional Maps', *IEEE Transactions on Neural Networks*, 3, 698-713.

6.2.5 Carpenter, G.A., Grossberg, S., and Rosen, D.B. (1991) 'Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system', *Neural Networks*, 4, 759-771.

6.2.6 Chen, M., Han, J. and Yu, P. (1996) 'Data mining: An overview from a database perspective', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, 866-883.

6.2.7 Freed, N. and Glover, F. (1981) 'Simple but powerful goal programming models for discriminant problems', *European Journal of Operational Research*, Vol. 7, 44-60.

6.2.8 Freed, N. and Glover, F. (1986) 'Evaluating alternative linear programming models to solve the two-group discriminant problem', *Decision Science*, Vol. 17, 151-162.

6.2.9 Glover, F. (1990) 'Improve linear programming models for discriminant analysis', *Decision Sciences*, Vol. 21, 771-785.

6.2.10 Guo, H. and Gelfand, S.B. (1992) 'Classification trees with neural network feature extraction', *IEEE Transactions on Neural Netwroks*, Vol. 3, 923-933.

6.2.11 Jang, J.-S. R., Sun, C.-T. and Mizutani, E. (1997), *Neuro-Fuzzy and Soft Computing*, Prentice Hall, Upper Saddle River, New Jersey.

6.2.12 Joachimsthaler, E.A. and Stam, A. (1988) 'Four approaches to the classification problem in discriminant analysis: An experimental study', *Decision Sciences*, Vol.19, 322-333.

6.2.13 Koehler, G. J. and Erenguc, S. S. (1990) 'Minimizing misclassifications in linear discriminant analysis', *Decision Science*, Vol. 21, 63-85.

6.2.14 Kosko, B. (1992), *Neural Networks And Fuzzy Systems: A Dynamical Systems Approach To Machine Intelligence*, Prentice Hall, Englewood Cliffs, New Jersey.

6.2.15 Markowski, E. P. and Markowski, C. A. (1985) 'Some difficulties and improvements in applying linear programming formulations to the discriminant problem', *Decision Science*, Vol. 16, 237-247.

6.2.16 Pass, S. (1997) 'Discovering in a value mountain of data', *ORMS Today*, October, 24-28.

6.2.17 Quinlan, J. (1986) 'Induction of decision trees', *Machine Learning*, Vol. 1, 81-106.

6.2.18 Rymon, R. (1993), Search through systematic set enumeration, Proceedings of the third international conference on principle of knowledge representation and reasoning. Cambridge, MA, 539-550.

6.2.19 Shi, Y., Y.J. Tian, G. Kou, Y. Peng, and J. P. Li (2009), *Optimization based Data Mining: Theory and Applications*, Springer.

6.2.20 Shi, Y, Peng, Y., Xu, X and Tang, X. (2002) 'Data mining via multiple criteria linear programming: Applications in credit card portfolio management', *International Journal of Information Technology and Decision Making*, Vol. 1, 145-166.

6.2.21 Wise, M., Luo, M. and Lin, Y. (2001) 'Data mining in credit card portfolio management: a multiple criteria decision making approach', in M. Koksalan and

S. Zionts, eds., *Multiple Criteria Decision Making in the New Millennium*, Springer, Berlin, 427-436.

6.2.22 Shi, Y. (1999) 'Data Mining', in *IEBM Handbook of Information Technology in Business*, Edited by M. Zeleny, International Thomson Publishing Europe.

6.3 Other sources of information.

6.4 Current bibliography of resource for student's information.

7.0 (IS Program) Estimate Computing Accreditation Commission (CAC) Category Content (class time in hours):

CAC Category	*Core*	*Advanced*
Hardware and software		
Networking and telecommunications		
Modern programming language		3
Analysis and Design		
Data management		36
Role of IS in Organizations		3

7.0 (CS Program) Estimate Computer Science Accreditation Board (CSAB) Category Content (class time in hours):

CSAB Category	*Core*	*Advanced*
Data structures		
Computer organization and architecture		
Algorithms and software design		24
Concepts of programming languages		

8.0 Oral and Written Communications:

Every student is required to submit at least ___9__ written reports (not including exams, tests, quizzes, or commented programs) to typically ___5__ pages and to make __1___ oral presentations of typically __10___ minutes duration. Include only material that is graded for grammar, spelling, style, and so forth, as well as for technical content, completeness, and accuracy.

9.0 Social and Ethical Issues:

No coverage

10.0 Theoretical content:

Please list the types of theoretical material covered, and estimate the time devoted to such coverage.

This course provides students theoretical issues for designing and constructing data warehouse and implementing data mining.

		Hours
10.1	The fundamental principles of Data Warehousing and Data Mining	5

11.0 Problem analysis:

Students learn theoretical and practical methods to design and construct data warehouse and implementing data mining. After covering the essential concepts, issues, techniques to build an effective data warehouse, this course emphasizes the various techniques of data mining, such as association, classification, clustering and prediction for on-line analyses within the framework of a data warehouse architectures. This course also promotes students to conduct a real-life data analyzing project.

12.0 Solution design:

Please describe the design experiences common to all course sections.

CHANGE HISTORY

Date	***Change***	***By whom***	***Comments***
10/7/03	Initial ABET version	Shi	
6/25/03	ABET cleanup	Wolcott	
10/09/07	ABET revision	Shi	

UNIVERSITY OF NEBRASKA AT OMAHA
Mapping of CS Program Outcomes vs. course objectives

Department and Course Number	
Course Title	
Course Coordinator	
Total Credits	
Date of Last Revision	

Instructions: Paste or type the course objectives in the left-hand column. Indicate the relationship between course objective and program outcome by placing one of the two following marks in the appropriate cell:

S – Strong relationship
X – Contributing relationship

Course objective	**CS Program Outcomes**										
	(a) knowledge of discipline	**(b) analyze problem, define**	**(c) design and implement solution**	**(d) function on a team**	**(e) ethical issues**	**(f) communicate effectively**	**(g) analyze impact of computing**	**(h) continued professional development**	**(i) Current techniques and tools**	**(j) apply foundations**	**(k) apply design and development**
1. Principles DW/DM	**X**	**X**									**X**
2. Transaction DB vs. DW		**X**									
3. OLAP/DW		**X**									
4. SQL			**S**						**S**		
5. Data mart			**S**						**S**		**X**
6. DW design/implementation			**S**	**S**					**X**		**S**
7. Data integration/transformation			**X**								**S**
8. DM algorithms		**X**	**S**							**S**	
9. Tools									**S**		
10. Interpret DM results					**X**	**S**	**S**				
11.											

CS Program Outcomes (2008)

(a) An ability to apply knowledge of computing and mathematics appropriate to the discipline;
(b) An ability to analyze a problem, and identify and define the computing requirements appropriate to its solution;
(c) An ability to design, implement and evaluate a computer-based system, process, component, or program to meet desired needs;
(d) An ability to function effectively on teams to accomplish a common goal;
(e) An understanding of professional, ethical, legal, security, and social issues and responsibilities
(f) An ability to communicate effectively with a range of audiences
(g) An ability to analyze the local and global impact of computing on individuals, organizations and society
(h) Recognition of the need for, and an ability to engage in, continuing professional development
(i) An ability to use current techniques, skills, and tools necessary for computing practices
(j) An ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices
(k) An ability to apply design and development principles in the construction of software systems of varying complexity.