Title (Units): **COMP7650 Data Mining and Knowledge Discovery (3,2,1)**

Course Aims: To introduce the fundamental issues of knowledge discovery and data mining; To learn the latest techniques of data mining and master the numeracy skills and quantitative methods for data mining; To conduct application case studies to show the usage of data mining for knowledge discovery.

Prerequisite: Basic Knowledge in Database Concepts, Probability and Statistics.

Course Intended Learning Outcomes (CILOs):
Upon successful completion of this course, students should be able to:

No.	Course Intended Learning Outcomes (CILOs)
	Knowledge
1	Identify and distinguish data mining applications from other IT applications
2	Explain data mining algorithms
3	Explain applicability of data mining applications
	Professional Skill
4	Suggest appropriate solutions to data mining problems
5	Analyze data mining algorithms and techniques
	Attitude
6	Develop team spirit in solving challenging data mining problems

Calendar Description: This course aims to introduce fundamental issues of knowledge discovery and the common data mining techniques including statistical methods, machine learning methods, and efficient algorithms. The main data mining techniques including classification, clustering, association rule mining, and also outlier detection. In addition, advanced topics such as graph mining and social network analysis will also be covered. Furthermore, their potential applications to a variety of areas such as business, finance, medicine, and so forth, are shown via some case studies.

Teaching and Learning Activities (TLAs):

CILOs	Type of TLA
1-3	Student will learn the concepts from lecture
4-5	Student will achieve the outcomes via assignment
4-5	Student will achieve the outcomes via guided laboratory
4-6	Student will achieve the outcomes via group project

Assessment:

No.	Assessment Methods	Weighting	CILOs to be addressed	Description of Assessment Tasks
1	Continuous Assessment	40%	4-6	Assignments and Labs will be used to consolidate their knowledge and develop their skills in data mining. Lab exercises will further strengthen their understanding and problem solving skills, which take 10% of an overall score. The written assignments and quizzes take 30% of an overall score.
2	Examination	60%	1-5	Final Examination questions are designed to evaluate how far students have achieved their intended learning outcomes. Analysis based questions will be used to assess the understanding of data mining problems. Problem solving questions will be used to assess the students' ability in tackling applications in data mining.

Assessment Rubrics:

	Excellent (A)	**Good (B)**	**Satisfactory (C)**	**Fail (F)**
Identify and distinguish data mining applications from other IT applications	Thorough identification of almost all applications	Identification of a large number of applications	Identification of a moderate number of applications	Identification of very small number of applications
Explain data mining algorithms	Thorough a detailed explanation of almost all data mining algorithms	Explanation of most of the algorithms	Explanation of some of the algorithms	Explanation of only a few number of algorithms
Explain applicability of data mining	Thorough a detailed explanation of almost all usage of data mining	Explanation of most of the usage	Explanation of some of the usage	Explanation of very small number of usage
Suggest appropriate solutions to data mining problems	Suggestion of almost all correct solutions	Suggestion of most of the solutions	Suggestion of some of the solutions	Suggestion of very small number of solutions
Analyze data mining algorithms and techniques	Thorough analysis of almost all data mining algorithms and techniques	Analysis of most of the data mining algorithms and techniques	Analysis of some data mining algorithms and techniques	Analysis of very small number of data mining algorithms and techniques

Course Content and CILOs Mapping:

Content		**CILO No.**
I	Overview of Knowledge Discovery and Data Mining	1
II	Foundations of Knowledge Discovery	2,3,5,6
III	Data Mining Techniques	2,3,5
IV	Case Studies of Data Mining in Social Network Analysis	3,4,5,6
V	Case Studies of Data Mining in Other Areas	3,4,5,6

References:

- P. Tan, M. Steinback and V. Kumar, Introduction to Data Mining, Pearson, Second Edition, 2018.
- J. Leskovec, A. Rajaraman, and J. Ullman, Mining of Massive Datasets, Second Edition, Cambridge University Press, 2014.
- J. W. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, CA, Third Edition, 2011.
- D. Hand, H. Mannila and P. Smyth, Principles of Data Mining, MIT Press, 2001.
- I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers, San Francisco, CA, Third Edition, 2011.

Course Content:

Topic

I. Overview of Knowledge Discovery and Data Mining

II. Foundations of Knowledge Discovery
   A. Data & similarity measures
   B. The process of knowledge discovery
   C. Different perspectives on knowledge discovery
   - A database perspective
   - A pattern recognition perspective
   - A statistical learning perspective

III. Data Mining Techniques
   A. Classification
   B. Clustering
   C. Association Rule Mining
   D. Outlier Detection
   E. Other related techniques

IV. Case Studies of Data Mining in Social Network Analysis

V. Case Studies of Data Mining in Other Areas