

Title (Units): SCI7570 DATA MINING AND KNOWLEDGE DISCOVERY (2,2,0)

Syllabus Reviewed by: Michael Ng

Prerequisite/Co-requisite: SCI7430 STATISTICAL SOFTWARE IN BUSINESS AND MANAGEMENT

Objectives: This subject introduces the concept of data mining and knowledge discovery, and data mining techniques (included advance statistical and machine learning technique) for solving problems such as data cleaning, clustering, classification, relation detection, forecasting.

Calendar Description: This subject introduces the concept of data mining and knowledge discovery, and data mining techniques (included advance statistical and machine learning technique) for solving problems such as data cleaning, clustering, classification, relation detection, forecasting.

References:

- 1) Han and M. Kamber, Data Mining: Concepts and Techniques, The Morgan Kaufmann Publishers, 2001.
- 2) P. Cabena, P. Hadjinian and R. Stadler, Discovery Data Mining From Concept to Implementation, Prentice Hall, 1997.
- 3) T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, 2001.
- 4) Westphal and T. Blaxton, Data Mining Solutions – Methods and Tools for Solving Real World Problems, John Wiley & Sons, 1998.
- 5) R. Kennedy, Y. Lee and R.P. Lippmann, Solving Data Mining Problems through Pattern Recognition, Prentice Hall, 1997.
- 6) A. Berson and S. Smith, Data Warehousing, Data Mining, & OLAP, McGraw-Hill, 1997.
- 7) Usama M. Fayyad et al, Advances in Knowledge Discovery and Data Mining, MIT Press, 1996.
- 8) Michalski Ryszard et al, Machine Learning and Data Mining Methods & Applications, John Wiley & Sons, 1998.
- 9) Berry, Michael J.A. and Gordon Linoff, Mastering Data Mining, John Wiley & Sons, 2000.

Assessment: Continuous assessment (35%) and Final examination (65%)

Subject Content in Outline:

- I. Introduction (4 hours of teaching)
 - A. The Knowledge Discovery Based in Databases (KDD)
 - B. Data and Data Visualization
 - C. Data Warehouse
 - D. Data Cleaning and Preprocessing
 - E. Data Mining Principles
- II. Mining Association Rules In Large Databases (6 hours of teaching)
 - A. Association Rule Mining
 - B. Mining Multidimensional Association Rules From Relational Databases

III. Statistical Tools In Data Mining (4 hours of teaching)

- A. Linear Regression
- B. Partial Regression
- C. Principal Components Analysis

IV. Classification, Forecasting And Decision Tree (14 hours of teaching)

- B. Similarity Measurement
- C. K – Means Clustering
- D. Hierarchical Clustering
- E. Decision Trees
- F. Machine Learning Techniques
- G. Classification and Regression Tree (CART)
- H. Boosting Methods
- I. Neural Networks