CITA/MINS 320

INTRODUCTION TO DATA MINING

Revised by: Eric Y. Cheng
A. **TITLE:** Introduction to Data Mining

B. **COURSE NUMBER:** CITA/MINS 320

**SHORT TITLE:** Data Mining

C. **CREDIT HOURS:** 3 credit hours

D. **WRITING INTENSIVE COURSE (OPTIONAL):** N/A

E. **COURSE LENGTH:** 15 weeks

F. **SEMESTER(S) OFFERED:** Spring

G. **HOURS OF LECTURE, LABORATORY, RECITATION, TUTORIAL, ACTIVITY:**
   3-hour lectures per week

H. **CATALOGUE DESCRIPTION:**
   This course is designed to provide a systematic introduction to the basic principles, methods, and applications of data mining. Students will gain knowledge on how data mining techniques work, how they can be applied across different domains by using these methods in real world. Topics include but are not limited to: decision trees, association rule discovery, clustering, classification, neural networks, and nearest neighbor analysis.

I. **PRE-REQUISITES/CO-COURSES:**
   Math 141 Statistics or permission of instructor

J. **GOALS (STUDENT LEARNING OUTCOMES):**

   By the end of this course, the student will be able to:

<table>
<thead>
<tr>
<th>Course Objective</th>
<th>Institutional SLO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Explain what tasks can be performed with Data Mining</td>
<td>2. Crit. Thinking</td>
</tr>
<tr>
<td>b. Explain data mining methodology and best practices</td>
<td>2. Crit. Thinking</td>
</tr>
<tr>
<td>c. Prepare data for data mining</td>
<td>2. Crit. Thinking</td>
</tr>
<tr>
<td>d. Apply Differential Responses Analysis method to analyze data and explain its outcomes</td>
<td>2. Crit. Thinking</td>
</tr>
<tr>
<td>e. Apply Chi-Square Test to analyze data and explain its outcomes</td>
<td>2. Crit. Thinking</td>
</tr>
<tr>
<td>f. Apply Decision Trees, Artificial Neural Networks, Nearest Neighbor Approaches, Market Basket Analysis, Automatic Cluster Detection methods to analyze data and explain its outcomes</td>
<td>2. Crit. Thinking</td>
</tr>
<tr>
<td>g. Evaluate and Determine when and where to use which data mining technique to analyze data</td>
<td>2. Crit. Thinking</td>
</tr>
</tbody>
</table>

K. **TEXTS:**
REFERENCES:

Tan, Pang-Ning, Steinbach, Michael, Kumar, Vipin, *Introduction to Data Mining*. Addison Wesley. 2005


M. EQUIPMENT: Technology Enhanced Classroom

N. GRADING METHOD: Standard A-F grading

O. MEASUREMENT CRITERIA/METHODS:

Assignments/report, team project, and tests.

P. DETAILED TOPICAL OUTLINE: See attached sheet

Q. LABORATORY OUTLINE: N/A
DETAILED TOPIC OUTLINE

MINS 320 Data Mining

TOPICS

I. Data and Data Mining
   A. Data Mining Tasks
   B. Data
      1. Types
      2. Data Quality
      3. Data Preprocessing
      4. Measures of Similarity and Dissimilarity
   C. Exploring Data
      1. Summary statistics
      2. Visualization
      3. OLAP and Multidimensional Data Analysis

II. Data Mining Techniques
   A. Classification
      1. Decision Tree Induction
      2. Model Overfitting
      3. Evaluating the Performance of a Classifier
      4. Methods for Comparing Classifiers
   B. Classification: Alternative Techniques
      1. Rule-Based Classifier
      2. Nearest-Neighbor Classifier
      3. Bayesian Classifier
      4. Artificial Neural Network
      5. Support Vector Machine
      6. Ensemble Methods
      7. Class Imbalance Problem
   C. Association Analysis
      1. Frequent Itemset Generation
      2. Rule Generation
      3. FP-Growth Algorithm
      4. Evaluation of Association Patterns
      5. Effect of Skewed Support Distribution
   D. Cluster Analysis
      1. K-Means
      2. Agglomerative Hierarchical Clustering
      3. DBSCAN
      4. Cluster Evaluation
E. Anomaly Detection
   1. Preliminaries
   2. Statistical Approaches
   3. Proximity-Based Outlier Detection
   4. Density-Based Outlier Detection
   5. Clustering-Based Techniques