MASTER SYLLABUS

CITA/MINS 320– INTRODUCTION TO DATA MINING

Created by: Eric Cheng
Updated by: Eric Cheng
A. % **TITLE:** Introduction to Data Mining

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

C. % **CREDIT HOURS:** (Hours of Lecture, Laboratory, Recitation, Tutorial, Activity)

  # Credit Hours: 3 !
  # Lecture Hours: 3 per week %
  # Lab Hours: per week %
  Other: per week

  Course Length: 15 Weeks

D. % **WRITING INTENSIVE COURSE:** No

E. % **GER CATEGORY:** No

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

G. % **COURSE 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.

H. % **PRE-REQUISITES/CO-REQUISITES:**

  a. Pre-requisite(s): MATH141 Statistics
  b. Co-requisite(s): None

I. % **STUDENT LEARNING OUTCOMES:**

<table>
<thead>
<tr>
<th>Course Student Learning Outcome [SLO]</th>
<th>PSLO</th>
<th>GER</th>
<th>ISLO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Explain what tasks can be performed with Data Mining</td>
<td>3. Demonstrate a solid understanding of the methodologies and foundations of IT</td>
<td>2 [CA]</td>
<td></td>
</tr>
<tr>
<td>b. Explain data mining methodology and best practices</td>
<td>3. Demonstrate a solid understanding of the methodologies and foundations of IT</td>
<td>2 [CA]</td>
<td></td>
</tr>
<tr>
<td>c. Prepare data for data mining</td>
<td>3. Demonstrate a solid understanding of the methodologies and foundations of IT</td>
<td>2 [CA] 5</td>
<td></td>
</tr>
<tr>
<td>d. Apply Differential Responses Analysis method to analyze data and explain its outcomes</td>
<td>3. Demonstrate a solid understanding of the methodologies and foundations of IT</td>
<td>2 [CA] 5</td>
<td></td>
</tr>
</tbody>
</table>
3. Demonstrate a solid understanding of the methodologies and foundations of IT

4. Apply problem solving and troubleshooting skills

5. Evaluate and Determine when and where to use which data mining technique to analyze data

**KEY**

<table>
<thead>
<tr>
<th>ISLO #</th>
<th>Institutional Student Learning Outcomes [ISLO 1–5]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Communication Skills</td>
</tr>
<tr>
<td></td>
<td>Oral [O], Written [W]</td>
</tr>
<tr>
<td>2</td>
<td>Critical Thinking</td>
</tr>
<tr>
<td></td>
<td>Critical Analysis [CA], Inquiry &amp; Analysis [IA], Problem Solving [PS]</td>
</tr>
<tr>
<td>3</td>
<td>Foundational Skills</td>
</tr>
<tr>
<td></td>
<td>Information Management [IM], Quantitative Lit./Reasoning [QTR]</td>
</tr>
<tr>
<td>4</td>
<td>Social Responsibility</td>
</tr>
<tr>
<td></td>
<td>Ethical Reasoning [ER], Global Learning [GL], Intercultural Knowledge [IK], Teamwork [T]</td>
</tr>
<tr>
<td>5</td>
<td>Industry, Professional, Discipline Specific Knowledge and Skills</td>
</tr>
</tbody>
</table>

**J. %**  APPLIED LEARNING COMPONENT: Yes _X_ No _____

**Classroom/Lab**

**K. %**  TEXTS:


L. REFERENCES:


M. EQUIPMENT:
Technology Enhanced Classroom

N. GRADING METHOD:
Standard A-F grading

O. SUGGESTED MEASUREMENT CRITERIA/METHODS:
Assignments, Exams, Project

P. DETAILED COURSE OUTLINE:

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

Q. LABORATORY OUTLINE: Not applicable