STATE UNIVERSITY OF NEW YORK %
COLLEGE OF TECHNOLOGY %
CANTON, NEW YORK %

MASTER SYLLABUS

CITA/MINS 320– INTRODUCTION TO DATA MINING

Created by: Eric Cheng
Updated by: Eric Cheng

Canino School of Engineering Technology %
Department of Decision & Graphic Media Systems %
Fall/2018 %

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:**

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

e. Apply Chi-Square Test to analyze data and explain its outcomes	3. Demonstrate a solid understanding of the methodologies and foundations of IT		2 [CA] 5
f. Apply Decision Trees, Artificial Neural Networks, Nearest Neighbor Approaches, Market Basket Analysis, Automatic Cluster Detection methods to analyze data and explain its outcomes	3. Demonstrate a solid understanding of the methodologies and foundations of IT 4. Apply problem solving and troubleshooting skills		2 [CA] 5
g. Evaluate and Determine when and where to use which data mining technique to analyze data	3. Demonstrate a solid understanding of the methodologies and foundations of IT 4. Apply problem solving and troubleshooting skills		2 [CA] 5

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

J. % APPLIED LEARNING COMPONENT: **Yes__X____** **No_____**

Classroom/Lab

K. % TEXTS:

Witten, Ian H. & Frank Eibe, Mark A. Hall, Chris J. Pal*, Data Mining: Practical Machine Learning Tools and Techniques,* 4th edition, Morgan Kaufmann, 2017. ISBN: 978-0128042915 / e-Book format is available / (Required: Data Mining Methods)

Berry, Michael J. A., Linoff Gordon S., *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 3rd edition, Wiley Computer Publishing, 2011. ISBN: 978-

0470650936 Companion web site: http://www.data-miners.com/dmt_companion.htm
(Recommended: Data Mining Applications)

Weka Documentation & Manual: https://www.cs.waikato.ac.nz/ml/weka/documentation.html

L. REFERENCES:

Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. 1st edition August 19, 2013, O'Reilly Media ISBN-13: 978-1449361327

Tan, Pang-Ning, *Introduction to Data Mining.* 2nd Edition 2018 Pearson Publishing ISBN: 9780133128901

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. <u>LABORATORY OUTLINE</u>: Not applicable