

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



**MASTER SYLLABUS**

**DATA 315 – DATA MINING AND MACHINE LEARNING**

**Created by: Kambiz Ghazinour**  
**Updated by:**

## SPRING 2023

- A. **TITLE:** Data Mining and Machine Learning
- B. **COURSE NUMBER:** DATA 315
- C. **CREDIT HOURS (Hours of Lecture, Laboratory, Recitation, Tutorial, Activity):**

# Credit Hours: 3  
# Lecture Hours per Week: 3  
# Lab Hours per Week:  
Other per Week:

Course Length (# of Weeks): 15 weeks (or 7 weeks)

- D. **WRITING INTENSIVE COURSE:** n/a
- E. **GER CATEGORY:** n/a
- F. **SEMESTER(S) OFFERED:** Fall and Spring

G. **COURSE DESCRIPTION:** This course provides the concepts and techniques in processing gathered data or information, which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This course focuses on the feasibility, usefulness, effectiveness, and scalability of techniques of large data sets. After describing data mining, this course explains the methods of knowing, preprocessing, processing, and warehousing data. Then, the methods involved in mining frequent patterns, associations, and correlations for large data sets are described. The course details the methods for data classification and introduces the concepts and methods for data clustering. Finally, it discusses the outlier detection and the trends, applications, and research frontiers in data mining.

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

a. Pre-requisite(s): None.

- I. **STUDENT LEARNING OUTCOMES:**

<b><u>Course Student Learning Outcome [SLO]</u></b>	<b><u>ISLO</u></b>
a. Examine the field of data science and familiarize with basics of data mining and machine learning.	2
b. Identify the process of collecting the data, pre-process and preparing it for data analysis.	2
c. Apply data classification and clustering methods on wide range of data to practice and	2

understand how the methods work.	
----------------------------------	--

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

J. **APPLIED LEARNING COMPONENT:** Yes \_\_\_\_\_ No **X** \_\_\_\_\_

If Yes, select one or more of the following categories:

Classroom/Lab \_\_\_\_\_  
 Internship \_\_\_\_\_  
 Clinical Practicum \_\_\_\_\_  
 Practicum \_\_\_\_\_  
 Service Learning \_\_\_\_\_  
 Community Service \_\_\_\_\_

Civic Engagement \_\_\_\_\_  
 Creative Works/Senior Project \_\_\_\_\_  
 Research \_\_\_\_\_  
 Entrepreneurship \_\_\_\_\_  
 (program, class, project)

**K. TEXTS:**

Data Mining: Concepts and Techniques  
3rd Edition by Jiawei Han, Micheline Kamber and Jian Pei

ISBN: 978-0-12-381479-1

ISBN10:0123814790

Copyright date: Jun 22, 2011

Pages: 744

**L. REFERENCES: n/a**

**M. EQUIPMENT: FLEX technology**

**N. GRADING METHOD: A-F**

**O. SUGGESTED MEASUREMENT CRITERIA/METHODS:**

- Participation
- Writing Assignments
- Case Study Projects

**P. DETAILED COURSE OUTLINE:**

- I. Introduction
  - A. What is Data Mining and what data can be mined?
  - B. What patterns can be detected
- II. Getting to know your data
  - A. Data Objects and Attribute Types
  - B. Basic Statistical Descriptions of Data
  - C. Data Visualization
- III. Data Preprocessing
  - A. Data Cleaning
  - B. Data Integration
  - C. Data Reduction
  - D. Data Discretization
- IV. Mining Frequent Patterns, Associations, and Correlations
  - A. Basic Concepts
  - B. Frequent Itemset Mining Methods
  - C. Which Patterns Are Interesting?—Pattern Evaluation Methods
- V. Classification
  - A. Decision Tree Induction
  - B. Bayes Classification Methods
  - C. Rule-Based Classification
  - D. Model Evaluation and Selection
- VI. Cluster Analysis: Basic Concepts and Methods
  - A. Cluster Analysis
  - B. Partitioning Methods
  - C. Hierarchical Methods
  - D. Density-Based Methods
  - E. Evaluation of Clustering

- VII. Outlier Detection
  - A. Outliers and Outlier Analysis
  - B. Outlier Detection Methods
  - C. Statistical Approaches
  - D. Proximity-Based Approaches

Q. **LABORATORY OUTLINE:**  
n/a