

CSE 5243: Introduction to Data Mining

Course Description

Knowledge discovery, data mining, data preprocessing, data transformations; clustering, classification, frequent pattern mining, anomaly detection, graph and network analysis; applications.

Prior Course Number: 674

Transcript Abbreviation: Intr Data Mining

Grading Plan: Letter Grade

Course Deliveries: Classroom

Course Levels: Undergrad, Graduate

Student Ranks: Senior, Masters, Doctoral

Course Offerings: Autumn, Spring

Flex Scheduled Course: Never

Course Frequency: Every Year

Course Length: 14 Week

Credits: 3.0

Repeatable: No

Time Distribution: 3.0 hr Lec

Expected out-of-class hours per week: 6.0

Graded Component: Lecture

Credit by Examination: No

Admission Condition: No

Off Campus: Never

Campus Locations: Columbus

Prerequisites and Co-requisites: Prereq: 3241 or 5241; and 2331, 5331, Stat 3301, or ISE 3200; and enrollment in CSE, CIS, ECE, Data Analytics or ISE major.

Exclusions:

Cross-Listings:

Course Rationale: Existing course.

The course is required for this unit's degrees, majors, and/or minors: No

The course is a GEC: No

The course is an elective (for this or other units) or is a service course for other units: Yes

Subject/CIP Code: 14.0901

Subsidy Level: Doctoral Course

Programs

Abbreviation	Description
BS CSE	BS Computer Science and Engineering
MS CSE	MS Computer Science and Engineering
PhD CSE	PhD Computer Science and Engineering

Course Goals

Be competent with anomaly detection algorithms and graph/network analysis algorithms
Master the knowledge discovery process
Be competent with simple data preprocessing and data transformation techniques

Master key classification and clustering algorithms
Master major frequent pattern mining algorithms

Course Topics

Topic	Lec	Rec	Lab	Cli	IS	Sem	FE	Wor
Knowledge Discovery Process and Background	3.0							
Elements of Data Preprocessing and Data Transformations	3.0							
Data Clustering	9.0							
Data Classification	6.0							
Frequent Pattern Mining	7.5							
Analyzing Graphs and Networks	7.5							
Anomaly Detection	3.0							
Applications (Bioinformatics, Social Networks)	3.0							

Representative Assignments

Data Preprocessing and Data Transformation Project/Assignment
Clustering and Classification Project/Assignment
Frequent Pattern Mining Project/Assignment
Network Analysis Project/Assignment

Grades

Aspect	Percent
Homeworks	20%
Midterm	20%
Final Exam	30%
Project/Programming Work	30%

Representative Textbooks and Other Course Materials

Title	Author
<i>Introduction to Data Mining</i>	Tan, Steinbach and Kumar, Addison Wesley, 2006
<i>Data Mining: Concepts and Techniques</i>	J. Han, M. Kamber : Morgan Kaufmann, 2006

ABET-EAC Criterion 3 Outcomes

Course Contribution		College Outcome
***	a	An ability to apply knowledge of mathematics, science, and engineering.
***	b	An ability to design and conduct experiments, as well as to analyze and interpret data.
***	c	An ability to design a system, component, or process to meet desired needs.
***	d	An ability to function on multi-disciplinary teams.
***	e	An ability to identify, formulate, and solve engineering problems.
*	f	An understanding of professional and ethical responsibility.
*	g	An ability to communicate effectively.

Course Contribution		College Outcome
*	h	The broad education necessary to understand the impact of engineering solutions in a global and societal context.
*	i	A recognition of the need for, and an ability to engage in life-long learning.
*	j	A knowledge of contemporary issues.
***	k	An ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

BS CSE Program Outcomes

Course Contribution		Program Outcome
***	a	an ability to apply knowledge of computing, mathematics including discrete mathematics as well as probability and statistics, science, and engineering;
***	b	an ability to design and conduct experiments, as well as to analyze and interpret data;
**	c	an ability to design, implement, and evaluate a software or a software/hardware system, component, or process to meet desired needs within realistic constraints such as memory, runtime efficiency, as well as appropriate constraints related to economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability considerations;
***	d	an ability to function on multi-disciplinary teams;
***	e	an ability to identify, formulate, and solve engineering problems;
*	f	an understanding of professional, ethical, legal, security and social issues and responsibilities;
*	g	an ability to communicate effectively with a range of audiences;
	h	an ability to analyze the local and global impact of computing on individuals, organizations, and society;
*	i	a recognition of the need for, and an ability to engage in life-long learning and continuing professional development;
*	j	a knowledge of contemporary issues;
***	k	an ability to use the techniques, skills, and modern engineering tools necessary for practice as a CSE professional;
**	l	an ability to analyze a problem, and identify and define the computing requirements appropriate to its solution;
***	m	an ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices;
**	n	an ability to apply design and development principles in the construction of software systems of varying complexity.

Prepared by: Mircea-Radu Teodorescu