IE Technical Elective – Data Mining

Fall 2024

Course Syllabus

(Sep 10 version, subject to change)

Instructor

Changxi Wang, Ph.D. (email: changxi.wang@scupi.cn)
Office: SCUPI Building N403
Office Hours: Wednesday 12:00 to 18:00

Teaching Assistant

Monarch Zhou 周昊鹏 (email: haopengzhou2000@163.com)
Office: SCUPI Building N407
QQ Group: 679140037

Lecture

Monday 13:50 to 16:25, Location: Zone 3, 104

Course Description

Data preprocessing, data visualization, classification, clustering, frequent patterns mining, association rules, Project presentation, Tableau and Python. 3 credit hours.

Course Prerequisites

IE 1070, MATH 280

Course Objectives

1. Learn basics of machine learning models.
2. Learn to identify industrial problems and formulate them into machine learning problems.
3. Learn to use programming languages to analyze data.
4. Learn to recognize and implement various ways of selecting suitable model parameters for different machine learning techniques.

Applicable ABET Outcomes

Students will build

1. An understanding of the data analytics lifecycle.
2. Skills in transformation and merging of data for use in analytic tools.
3. An overview of simple statistical models and the basics of machine learning techniques such as clustering, associations, classification.
4. An understanding of good practices of data science, and conversely recognizing bad practices and why.
5. Skills in the use of tools such as Python, Tableau to explore and mine simple data sets.

Textbook

Jiawei Han, Micheline Kamber, Jian Pei, Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems) 3rd Edition, 2012

References

数据挖掘导论 Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Introduction to data mining. Pearson Education India, 2016.

Grading

Homework & exercises, projects, and exam questions related specifically to the objectives above.

Attendance, Homework, Exercises & Quizzes:	30%
Project:	40%
Final Examination:	30%
	100%

Score	Letter Grade
90.00-100.00	A
85.00-89.99	A-
80.00-84.99	B+
76.00-79.99	B
73.00-75.99	B-
70.00-72.99	C+
66.00-69.99	C
63.00-65.99	C-
61.00-62.99	D+
60.00-60.99	D
0.00-59.99	F

Attendance

Attendance will be taken for each lecture period.

Homework & Exercises

Homework will be assigned weekly and needed to be finished before the next class. Homework solutions must be submitted to the Blackboard system.

Exams

There will be one open book open notes exam.

Group Project

Group project will be described in separate handouts as they are assigned.

Avoiding Plagiarism

1. Unacknowledged direct copying from the work of another person, or the close paraphrasing of somebody else's work, is called plagiarism and is a serious offence, equated with cheating in examinations. This applies to copying both from other students' work and from published sources such as books, reports or journal articles.
2. Paraphrasing, when the original statement is still identifiable and has no acknowledgement, is plagiarism. A close paraphrase of another person's work must have an acknowledgement to the source. It is not acceptable for you to put together unacknowledged passages from the same or from different sources linking these together with a few words or sentences of your own and changing a few words from the original text: this is regarded as over-dependence on other sources, which is a form of plagiarism.

SCUPI
Sichuan University - Pittsburgh Institute

Tentative Course Schedule

Week	Dates	Topics	Chapter
1	Sep 2	Course Introduction and Review of Syllabus, Introduction to Data Analytics and Data Mining	1
2	Sep 9	Know your data	2
3	Sep 14	Visualization and Distance Measures	2
4	Sep 23	Introduction to Tableau visualization software, Introduction to Python Programming I	
5	Sep 30	Data Preprocessing	3
6	Oct 12	Data Preprocessing: PCA & Feature Selection, Introduction to Python Programming II	6
7	Oct 14	Mining Frequent patterns and Associations with Apriori algorithm; Frequent Pattern Growth Method	7
8	Oct 21	Classification I: introduction to decision trees, Bayes Naïve, KNN / Midterm Exam Review	8
9	Oct 28	Classification II: Rule Based, Project Proposal Due	8
10	Nov 4	Classification III: logistic Regression, SVM	8
11	Nov 11	Classification III: Evaluating Accuracy	8
12	Nov 18	Python Programming III	
13	Nov 25	Clustering I	10
14	Dec 2	Clustering II	10
15	Dec 9	Final Exam Review	
16	Dec 16	Project Presentation	
17	Dec 23	Final Exam	