IE Technical Elective – Data Mining
Fall 2022
Course Syllabus
(Sep 3 version, subject to change)

Instructor
Changxi Wang, Ph.D. (email: changxi.wang@scupi.cn)
    Office: Zone 4-220
    Office Hours: Monday 12:20 to 13:40, Wednesday 15:00-16:40 / Tencent Meeting: TBD

Teaching Assistant
Monarch Zhou 周昊鹏 (email:2018141523010@stu.scu.edu.cn)
QQ Group: 659736183

Lecture
Monday 13:50-16:25; Location: Teaching Building 1, D203 / Tencent Meeting: TBD

Course Description
Data preprocessing, data visualization, classification, clustering, frequent patterns mining, association rules, regression. Project presentation. 3 credit hours.

Course Prerequisites
IE 1070, MATH 280

Course Objectives
1. Learn basics of machine learning models used in industrial engineering.
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:	20%
Project:	20%
Mid-Semester Examination:	25%
Final Examination:	<u>35%</u>
	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 two exams, a formula sheet will be provided for each exam.

<u>Group Project</u>

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.

Tentative Course Schedule

Lecture	Dates	Topics	Chapter
1	Sep 5	Course Introduction and Review of Syllabus, Introduction to Data Analytics and Data Mining	1
2	Sep 12	Know your data	2
3	Sep 19	Visualization and Distance Measures	2
4	Sep 26	Data Preprocessing Introduction	3
5	Oct 8	Data Preprocessing: PCA & Feature Selection Review Mining Frequent patterns and Associations with Apriori algorithm	6
6	Oct 10	More Frequent patterns and Associations – Apriori continued; Frequent Pattern Growth Method	7
7	Oct 17	Classification I: introduction to decision trees, Bayes Naïve, KNN / Midterm Exam Review	8
	Oct 24	Midterm Exam	
8	Oct 31	Project Proposal Due today, Tableau visualization software,	
9	Nov 7	Python Programming I	
10	Nov 14	Classification II: Rule Based, logistic Regression, Support Vector Machine	8
11	Nov 21	Classification III: Evaluating Accuracy	8
12	Nov 28	Regression	
13	Dec 5	Clustering I	10
14	Dec 12	Clustering II, Final Exam Review	10
15	Dec 19	Project Presentation	
16	Dec 26	Final Exam	