

SCIM 323, SCMA 447 – Data Mining Syllabus and Timetable

Instructor: Dr. Pannapa Changpetch (pannapa.cha@mahidol.edu)

Office hours: By appointment

Text and Electronic Resources:

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., Lichtendahl Jr., K. C. (2018) Data Mining for Business Analytics: Concepts, Techniques, and Applications in R 1st Edition, Wiley.

The text will be supplemented with materials prepared by the instructor and additional readings. Google drive will hold copies of daily handouts and materials.

Software: R (available at <http://www.r-project.org>)

Learning Objectives: The use of analytic models to improve business performance is a large and expanding business area. Creative problem solvers in this area are in high demand in the marketplace, and entrepreneurial thinkers have opportunities in all kinds of organizations, large and small, public and private, for profit and not for profit.

Course descriptions

Introduction to data mining; virtuous cycle of data mining; data mining methodology and best practices; data for data mining; introduction to classification: Naive-Bayes and nearest neighbor; decision trees; estimating the predictive accuracy of a classifier; continuous attributes; avoiding overfitting of decision trees; association rule mining; clustering.

Grading:	Attendance, participation and class activities	10%
	Homework Assignments	7%
	Computer Exam	16%
	Project	30%
	Final Exam*	37%
	Total	100%

Note*: In case that we cannot have an on-site final exam, we will replace the 37% of this category by the second project and the second computer exam instead.

Tentative Timetable

Week	Topics
1	Introduction to data mining
2	Data visualization
3-4	kNN
5-6	Decision tree
7-8	Naïve Bayes
9	Midterm exam
10	Association rules
11-12	Cluster analysis
13-14	Presentation part 1
15	Topics in time series
16	Additional topic / Case study
17	Final exam

Special Notes:

1. Topics may be moved from week to week depending on how long coverage of each topic takes.
2. You are responsible for checking all the LINE messages I will send during the semester.
3. All homework assignments must be submitted by the due time on the Google Classroom. Late homework will not be accepted.
4. Project submitted up to 24 hours late will lose 15%. Late project after 24 hours will not be accepted.

Participation Expectations: Students should be in class on time and be ready to participate. Students should plan on actively engaging in class discussions by providing constructive comments. Note that course materials and lecture notes will be posted on the Google Drive.

Personal technology: Please do not use a personal device for any purpose unrelated to our class.

Course Structure: Daily classes will be interactive with problem solving emphasized.

Homework: The homework is meant to be an individual exercise. Therefore, it must be **a student's own work in its entirety**. This includes coding, results and all calculations. If two students submit homework problems that have identical and highly unlikely errors, this is evidence that the students did not work the problems individually.

Exam: In-class exam must be completed individually without the help of prohibited materials. All electronic devices except a calculator are not to be used under any circumstances.

Project: The work in the project must be original. For the team project, you share responsibility for the work of the entire team. If plagiarism or the use of prohibited material is detected in the work of one team member, the entire team will be sanctioned.

Grading system

Score			Grade
85	-	100	A
75	-	84.99	B+
70	-	74.99	B
65	-	69.99	C+
60	-	64.99	C
55	-	59.99	D+
50	-	54.99	D
0	-	49.99	F