Data Mining

13016366

Dr.Pimprapai Thainiam

International College

▶ Contact Information

Instructor : Dr. Pimprapai Thainiam (Wow)

Office : IC07 Logistics and Supply Chain Management on 8th floor

Email : pimprapai.th@kmitl.ac.th

Office Hours : Monday 1.00 pm – 4.00 pm or by appointment via email

International College

▸ About the Course

Class   : Tuesday 9.00 am – 12.00 pm at IC03

First Class   : Tuesday, August 7th, 2018

Midterm Exam : Tuesday, September 25th, 2018

Last Class   : Tuesday, November 27th, 2018

Final Exam   : Thursday, December 7th, 2018

Course Grade : December 18th, 2018

Prerequisites   : 13006009 - Probability and Statistics

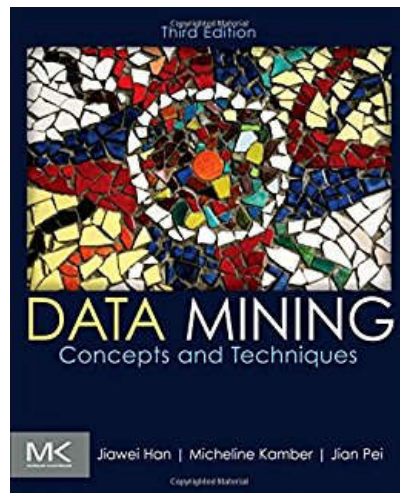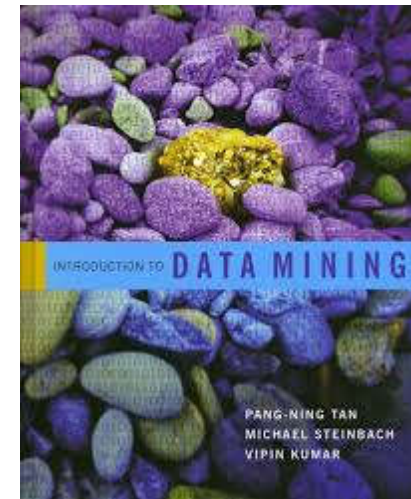   13016237 - Information Systems and Databases

▸ **Course Description**

This course is an introduction to fundamentals of data mining. Students will learn the basics of data mining algorithms with an emphasis on their real-world applications. The course will cover the basic concepts of data mining; data types; performance evaluation; data mining techniques:

- Classification

- Association rules mining

- Clustering

- Regression

International College

▶ Course Textbooks

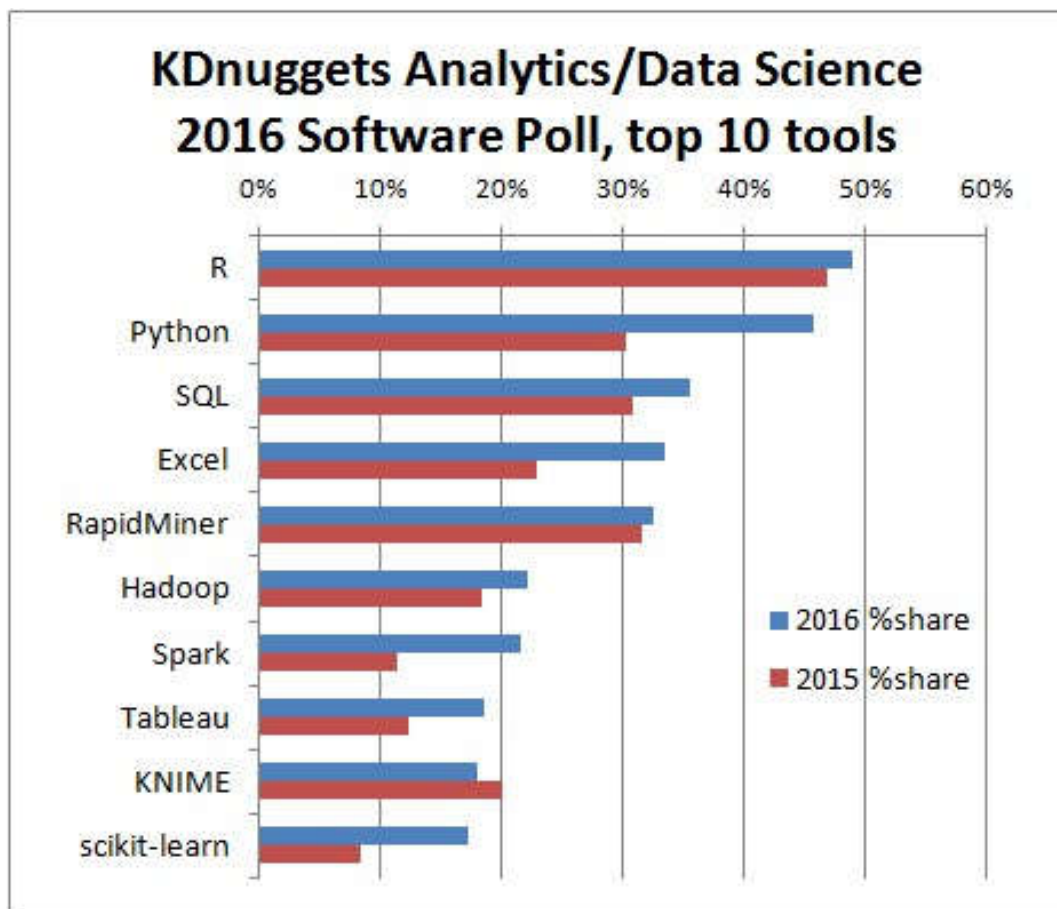**Introduction to Data Mining**

by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar

Data mining: Concepts and Techniques

by Jiawei Han and Micheline Kamber

▶ Programming Language

What software you used for Analytics, Data Mining, Data Science, Machine Learning projects in the past 12 months?
2,895 voters and 102 different tools

Next table has the top 10 most popular tools in 2016 poll

Tool	2016 % share	% change	% alone
R	49%	+4.5%	1.4%
Python	45.8%	+51%	0.1%
SQL	35.5%	+15%	0%
Excel	33.6%	+47%	0.2%
RapidMiner	32.6%	+3.5%	11.7%
Hadoop	22.1%	+20%	0%
Spark	21.6%	+91%	0.2%
Tableau	18.5%	+49%	0.2%
KNIME	18.0%	-10%	4.4%
scikit-learn	17.2%	+107%	0%

https://www.kdnuggets.com/2016/06/r-python-top-analytics-data-mining-data-science-software.html

▸ **Programming Language**

R : https://cran.r-project.org/

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

R Studio : https://www.rstudio.com/products/rstudio/download/

	rstudio::conf	Products	Resources	Pricing	About Us	Blogs	Q
License	AGPL	Commercial	AGPL	Commercial		Commercial	
Pricing	FREE	$995/yr	FREE	$9,995/yr		$29,995/yr	
	DOWNLOAD	BUY NOW	DOWNLOAD	DOWNLOAD		CONTACT SALES	
	Learn More	Learn More	Learn More	Learn More		Learn More about RStudio Connect	

International College

Evaluation and Grading

Assignment 1 (Data Exploration, Cleaning, and Visualization)	10%
Assignment 2 (Classification)	10%
Assignment 3 (Association Rules Mining)	10%
Assignment 4 (Clustering)	10%
Midterm Exam	30%
Final Exam	30%

Final grade evaluation will be a subjective process; however, it will be based primarily on your understanding of the material as evidenced in reports, midterm and final exams.

Assignment and Exam Details

- Group of 2 people

- Each assignment must be a detailed assignment report

- No limitation on number of pages

- Please submit your assignment reports as pdf files via pimprapai.th@kmitl.ac.th by 9.00 am on their assigned due dates.

- Example of how to name your files: Name 1_Name 2_Project 1.pdf

- Late work will be penalized 1 point per day (24-hour period) (Each assignment has 10 points). This penalty will apply except in case of documented emergency (e.g., medical emergency)

- Midterm and final exams will be open notes/open books.