COMPUTER SCIENCE

INDIANA UNIVERSITY

School of Informatics and Computing
Bloomington

DATA MINING

CSCI-B565

Spring 2018

PROFESSOR

Class meets:

Time: TR 11:15pm – 12:30pm
Place: Fine Arts 102

Instructor:

Predrag Radivojac
Office: Luddy Hall 2048
Email: predrag@indiana.edu
Web: www.cs.indiana.edu/~predrag

Office Hours:

Time: Tuesdays and Thursdays 2:00pm-3:00pm
Place: Luddy Hall 2048

Course Web Site:

https://www.cs.indiana.edu/~predrag/classes/2018springb565/

Associate Instructors aka Teaching Assistants

Moses Stamboulian

Email: mstambou

*Office hours: MW 4-5:30

Eriya Terada

Email: eterada

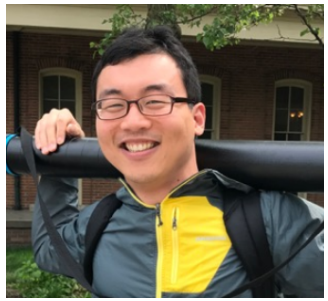
*Office hours: TR 9:30-11

Benjamin Rosenzweig

Email: bkrosenz

*Office hours: TR 3-4:30

Additional Associate Instructors



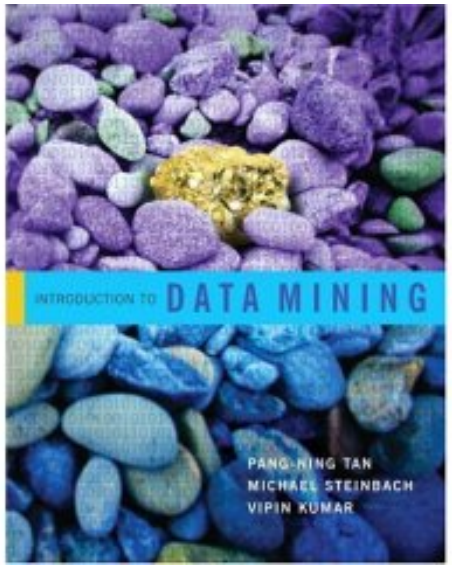
Yuxiang Jiang

Email: yuxjiang

*Office hours: MW 9:30-11

TEXTBOOK

Introduction to Data Mining - by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar

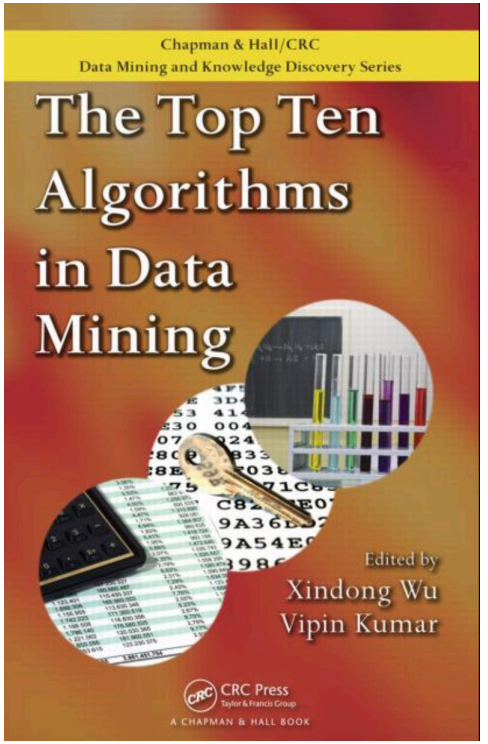


Chapter 1: Introduction
Chapter 2: Data
Chapter 3: Exploring data
Chapter 4: Classification: Basic
Chapter 5: Classification: Advanced
Chapter 6: Association analysis: Basic
Chapter 7: Association analysis: Advanced
Chapter 8: Cluster analysis: Basic
Chapter 9: Cluster analysis: Advanced
Chapter 10: Anomaly detection

Supplementary material will be provided in class!

SECOND TEXTBOOK

The Top Ten Algorithms in Data Mining - by Xindong Wu and Vipin Kumar

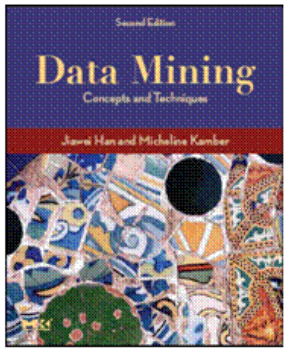


Chapter 1: C4.5
Chapter 2: K-means
Chapter 3: SVM: Support Vector Machines
Chapter 4: Apriori
Chapter 5: EM
Chapter 6: PageRank
Chapter 7: AdaBoost
Chapter 8: kNN: k-Nearest Neighbors
Chapter 9: Naïve Bayes
Chapter 10: CART: Classification and Regression Trees

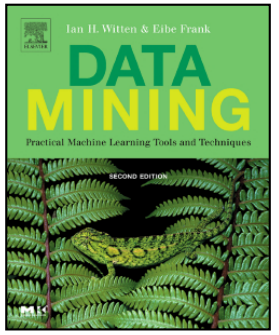
Supplementary material will be provided in class!

ALSO GOOD READINGS...

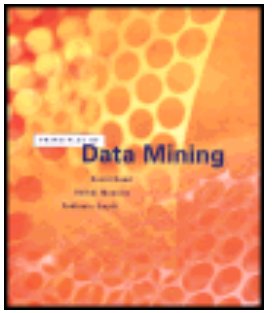
- Data Mining: Concepts and Techniques - by Jiawei Han and Micheline Kamber



- Data Mining: Practical Machine Learning Tools and Techniques - by Ian Witten and Eibe Frank

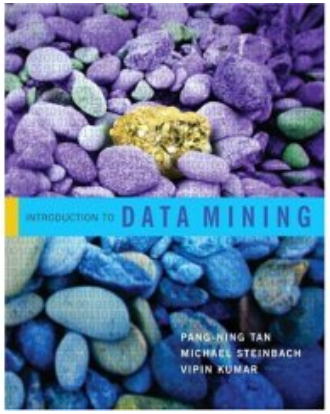


- Principles of Data Mining - by David Hand, Heikki Manilla, and Padhraic Smyth

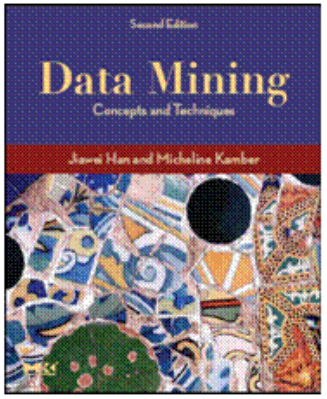


LECTURE SLIDES

- Introduction to Data Mining - by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar



- Data Mining: Concepts and Techniques - by Jiawei Han and Micheline Kamber



- Summary: Our own slides + some mix from the slides for the books above

Main Definitions and Goals

- Data mining is a well-founded practical discipline that aims to identify interesting new relationships and patterns from data (but it is broader than that).

- This course is designed to introduce basic and advanced concepts of data mining and provide hands-on experience to data *analysis*, *clustering*, and *prediction*.

- The students are be expected to develop a working understanding of data mining and develop skills to solve practical problems.

JUNE 22, 2015

TIME

How High Is Your XQ?

Your next job might depend on it

BY ELIZA GRAY

Is it true to say you have never hated anyone? Do you understand why stars twinkle? Have you used a display of emotion to get what you want? Would you rather read or watch TV? Do you usually notice when you are boring people? Do you hate opera singing? Would you co[illegible]er yourself to be an ordinary person? Are you shy? Do you prefer problems t[illegible]require a lot of thought? Do you enjoy giving parties? When yo[illegible]frequently rebellious? Do you prefer to work with [illegible]o you believe people get stressed when they try t[illegible]mething in your day that makes you feel happy? H[illegible]u uncomfortable accepting help from ot[illegible]ever stressed at work? Do you think sometim[illegible]le? Do you like to have someone around at wo[illegible]uld you like to change a lot of things about yo[illegible]ess? Do you make new friends all the time? D[illegible]sh you? Have you ever pretended to know m[illegible] to form friendships at work? Would your colleagu[illegible]ou are ver[illegible]confident? How much does

Do you often fantasize about being famous?

Do you find yourself getting angry easily?

Would you like to be an art collector?

Do people say you are eccentric?

Find out if your personality fits your job

For a better sense of what these tests are like, TIME asked Hogan Assessments to devise a brief example for readers. The company's co-founder, Robert Hogan, is a fellow at the Society for Industrial and Organizational Psychology. He says he believes testing can strengthen organizations and place people in the right jobs—but adds that he's concerned about the lack of transparency and regulation in the growing test industry.

Instructions
Answer each of the following true/false statements. To get your score, give yourself a point each time you answered true to the following items: 1, 3, 6, 7 and 8. Give yourself a point each time you answered false to the remaining items (2, 4 and 5).

Spontaneity

T F 1. I follow my instincts wherever they lead me.
T F 2. Planning is one of my best abilities.
T F 3. My friends would describe me as impulsive.
T F 4. I always schedule my activities well in advance.
T F 5. I don't do anything without having a plan in place.
T F 6. I often do things on the spur of the moment.
T F 7. I go wherever the day takes me.
T F 8. My friends say that I am unpredictable.

Now add up your total score.

_____ **Score**

High (6–8):
You would fit best in jobs that offer flexibility and diversity. Others may turn to you for your spontaneity and willingness to change focus, but you might have difficulty with details and completing everything that you start.

Moderate (3–5):
You would fit best in jobs that offer some structure but still permit a change of focus and direction throughout the day. Others will view your ability to plan ahead while remaining flexible as an asset, but you might have difficulty prioritizing your work or quickly bouncing from one task to another.

Low (0–2):
You would fit best in a structured working environment where you can plan your day in advance and stick to that plan. Others will likely rely on you for your attention to detail and follow-through, but you might have difficulty recognizing when it is time to abandon one course of action for another.

Independence

T F 1. I prefer working alone.
T F 2. I am a social person.
T F 3. I don't like to rely on others to help me with my job.
T F 4. I usually spend my free time with friends.
T F 5. I enjoy working on teams.
T F 6. I like to work without distractions.
T F 7. Meeting with others is often a waste of time.
T F 8. I do not like to draw attention to myself.

Now add up your total score.

_____ **Score**

High (6–8):
You would fit best in jobs that allow you to work primarily by yourself. Others will likely value your ability to take on and run with your own projects, but you might have difficulties asking for or lending help to others when needed.

Moderate (3–5):
You would fit best in jobs that contain a mix of independent work and working with others. Others will value your flexibility, but you might struggle during long periods of solitary work or work that requires constant and regular interactions with others.

Low (0–2):
You would fit best in an environment where you are constantly surrounded by and working with others. You likely derive energy from others, which in turn attracts people to you. You might struggle, however, when asked to focus on specific tasks or goals by yourself.

Competitiveness

T F 1. Life is a competition.
T F 2. I don't care if others are more successful than I am.
T F 3. Some people think I am too competitive.
T F 4. There's nothing wrong with letting others win.
T F 5. More harm than good is caused by competition.
T F 6. It takes a killer instinct to get ahead.
T F 7. I enjoy testing my skills against others.
T F 8. I am at my best when competing with others.

Now add up your total score.

_____ **Score**

High (6–8):
You would fit best in jobs where success is defined by excelling over others. Others will likely value your drive, determination and willingness to put in more effort than your competition. You might, however, struggle to recognize when you need to put your own personal goals aside for the greater good.

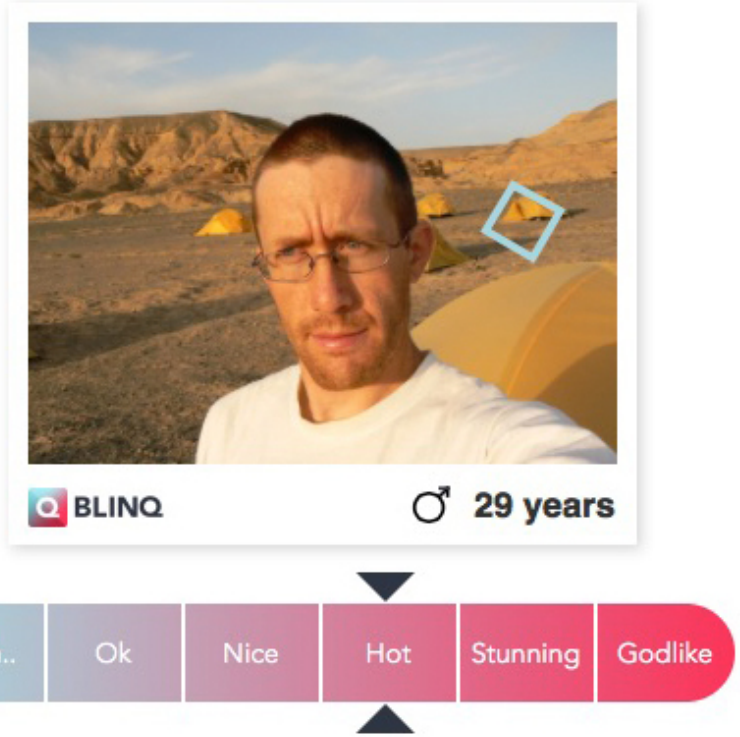
Moderate (3–5):
You would fit best in environments that contain a mix of collaboration and competition. You are likely able to work toward both individual and team goals simultaneously but must be careful to recognize when to focus on one over the other.

Low (0–2):
You would fit best in a team-based, collaborative environment. Others will frequently turn to you for help and, in turn, be willing to offer you help as needed. You might, however, struggle in situations where your personal performance depends on your ability to outshine others.

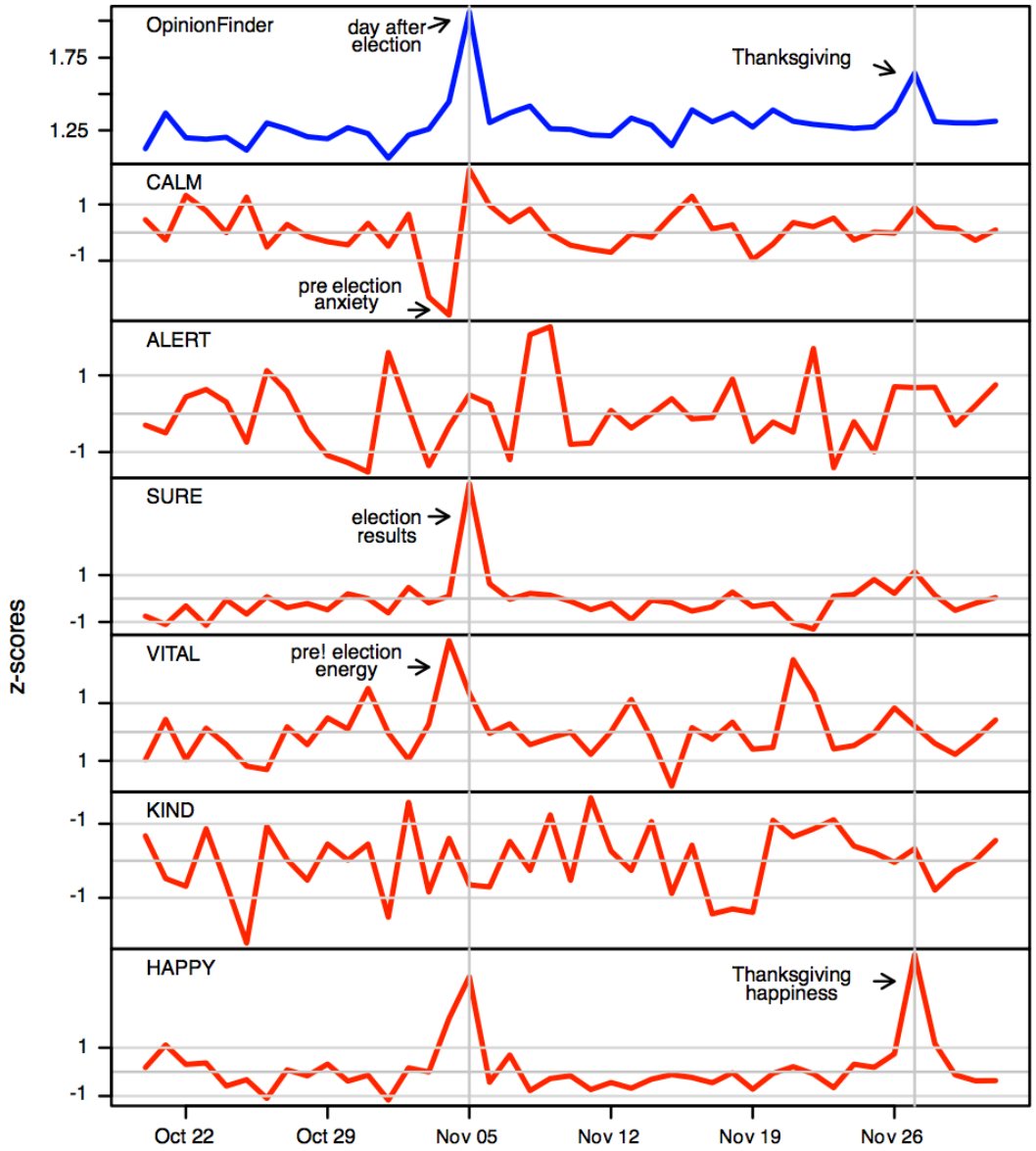
Motivating Example #2, from Reddit

Let Artificial Intelligence guess your attractiveness and age

#howhot



TWITTER VS. STOCK MARKET



* Bollen et al. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 2011, Pages 1-8.

EXPECTATIONS AND ASSUMPTIONS

- Basic mathematical skills
  - calculus, probabilities, linear algebra
- Good programming skills
  - programming languages: Matlab, Python, R, C/C++
- You are patient and hardworking
- You are motivated to learn and succeed in class
- Your integrity is impeccable

GRADING: BASIC

- Midterm exam (w9): 20%
- Final exam (May 3, 12:30-2:30pm): 20%
- Homework assignments (4): 35%
- Class project: 20%
- Class participation ~~and quizzes~~: 5%

GRADING: ADVANCED

- Top performers in the class will get As
- Distributions of scores will be shown*
- If you don't know where you stand in class, ask me*
- All assignments count, must be typed to show formulas properly! Plan ahead!
- All assignments are individual!
- All the sources used for problem solution must be acknowledged (people, web sites, books, etc.)

* not before the midterm exam is graded

LATE ASSIGNMENT POLICY

- The homework assignments are due on the specified due date through Oncourse

- Late assignments will be accepted* using the following rules

– **points**	**(on time)**	}	**recommended!**
– points x 0.9	(1 day late)	}	not recommended!
– points x 0.7	(2 days late)		
– points x 0.5	(3 days late)		
– points x 0.3	(4 days late)		
– points x 0.1	(5 days late)		
– 0	(after 5 days)		

* if there are legitimate circumstances to not apply this policy, please inform me early

ROADMAP

January:	8	March:	5
	15		12 (SB)
	22		19
	29		26
February:	5	April:	2
	12		9
	19		16
	26		23
			30*

ROADMAP

January:	8	March:	5 M
	15 h1		12 (SB)
	22		19 h3
	29 H1		26
February:	5 h2	April:	2 H3
	12		9 h4
	19 H2, pp		16 H4
	26 PP		23 P
			30* F

ACADEMIC HONESTY

- *Code of Student Rights, Responsibilities, and Conduct* !!!
  - http://studentcode.iu.edu/
  - Many interesting things there, including that… Students are responsible to "facilitate the learning environment and the process of learning, including attending class regularly, completing class assignments, and coming to class prepared".

- Academic honesty taken seriously!
  - I am obliged to report every cheating incident to the university
  - Do the right thing

MISCELLANEA

- Do not record instructor(s) without written permission
  - covers: lectures and office hours

- Turn off cell phones and other similar devices during class

- Use laptops if you have to (unless it bothers someone)
- “will u be in ur office after class”; “I need a letter of recommendation.”
- BE NICE TO PEOPLE