

Rutgers University
The State University of New Jersey
Department of Economics - CCAS
Spring 2018

Class Information

Course Title: Applied Data Mining & Knowledge Discovery (index#16064)
Economics 366/Section 01

Instructor: Dr. I-Ming Chiu

Office: ARMITAGE 328
Phone (856) 225 6012

E-mail address: ichiu@camden.rutgers.edu

Class Meeting: BSB 335. 11:10 AM-12:30 PM (Tuesday & Thursday)

Office Hours: 3:30-4:30 pm, Tuesday & Thursday or by appointment

Course Description: We are living in an era where new data are being fast produced from all kinds of human activities that include our daily emails, social network postings, and transactions on S&P 500 etc. How do we utilize these high-volume data? The answer is “Data Mining”. Data mining is a computing process of using mathematical and statistical algorithms to uncover patterns from huge data and convert them into decision-making information and new knowledge. A variety of data mining algorithms will be introduced and the topics include linear regression model, logistic regression model, KNN, Support Vector Machine, Naïve Bayes, Association Rules, and K-Means Clustering, etc. Each topic covered is accompanied with a case study, so students will gain many hands-on learning experiences. The ultimate goal in this course is to equip students with modern data analytical tools, which has a high demand in the job market.

Required Reading: Abhijit Ghatak, Machine Learning with R, Springer, 2017.
N.D. Lewis, Machine Learning Made Easy with R: An Intuitive Step by Step Blueprint for Beginners, CreateSpace Independent Publishing Platform, 2017.
(Electronic edition can be purchased at Amazon.com)

Both of the above required books are available for purchase at the University District Bookstore (601 Cooper St., Camden, NJ 08102)

Other References: Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, An Introduction to Statistical Learning/with Applications in R, Springer, 2013.

Taylor Arnold & Lauren Tilton, Humanities Data in R/Exploring Networks, Geospatial Data, Images, and Text, Springer, 2015.

Rutgers University
The State University of New Jersey
Department of Economics - CCAS
Spring 2018

Max Kuhn & Kjell Johnson, Applied Predictive Modeling, Springer, 2013.

Norman Matloff, Statistical Regression and Classification/From Linear Models to Machine Learning, CRC Press, 2017.

Johannes Ledolter, Data Mining and Business Analytics with R, Wiley, 2013.

Song Yang, Franziska B. Keller, and Lu Zheng, Social Network Analysis/Methods and Examples, Sage 2017.

Computing: All the computations will be done using an open source statistical software R. It can be downloaded at <http://www.r-project.org>. You're encouraged to download and use RStudio at the following site, which is a very user-friendly IDE (integrated development environment) for R. <https://www.rstudio.com/products/rstudio/>

R Installation: <https://www.youtube.com/watch?v=Icawuhf0Yqo> (for Mac)
<https://www.youtube.com/watch?v=hxj0UG4boGU> (for PC)

Class Material: Data, handouts, assignments, and additional readings will be posted on Sakai website: <https://sakai.rutgers.edu/portal>.

Online Learning: <https://www.datacamp.com/> (Learn Data Science online)

Useful Websites: <http://www.statmethods.net/> (Computing using R web site)
<http://www.ats.ucla.edu/stat/> (Computing learning at UCLA)
<http://socserv.mcmaster.ca/jfox/> (Dr. Fox's statistics site)
<https://www.analyticsvidhya.com/> (Machine Learning)

Spring '18 Calendar: <http://registrar.camden.rutgers.edu/academic-calendar-2017-2018>

Grading: **Contribution to Final Grade**

- Attendance	5%
- Take-home problems	30%
- Midterm Exam (2)	40%
- Final Exam/Project	25%
- Participation (extra credit)	5%

Grading Policy: Term grades will be based on the final distribution of the above grading weights.

Rutgers University
The State University of New Jersey
Department of Economics - CCAS
Spring 2018

- Exam Preparation:** The exam questions will be drawn from three sources: (i) homework assignments, (ii) course lectures, and (iii) reading material.
- Class Participation:** Class attendance is essential for learning achievement. When missing a class, it would cost you more time to learn on your own. I strongly recommend the following steps for your successful learning: (1) attend every class and take notes; (2) review everything you learn from the class immediately, never put it off; (3) ask questions and participate in class discussions.
- Academic conduct:** Make up exams will be given **only upon prior notice**. I request prior knowledge of any expected absence from an exam. If this is not feasible, you can document a valid reason for missing the exam. Unexcused absence on any exam will result in a grade of zero. Dishonesty in seeking an excused absence or in the examination process will result in a grade of zero on the exam involved and in university discipline. More detailed information can be found at the following site: <https://academicintegrity.rutgers.edu>.
- Disability Services:** Students with disabilities should contact the Rutgers-Camden Office of Disability Services (ODS). For more information, visit <http://learn.camden.rutgers.edu/disability-services>. Accommodation will be made in accordance with Rutgers University Policy.

Rutgers University
The State University of New Jersey
Department of Economics - CCAS
Spring 2018

Course Outline:

Topic 1	Mathematical and Statistical Fundamentals
Topic 2	Introduction to Computing Using R
Topic 3	What is Data Mining? The Familiar Linear Regression Model
Topic 4	K Nearest Neighbors (KNN)
Topic 5	Naïve Bayes
Exam 1	Date: TBA in the class
Topic 6	Decision Trees
Topic 7	Linear Discriminant Analysis
Topic 8	Logistic Regression Model
Topic 9	Support Vector Machine
Exam 2	Date: TBA in the class
Topic 10	Association Rules
Topic 11	Social Network Analysis
Topic 12	K-Means Clustering
Final Exam (or Project)	11:30 AM~2:30 PM, Tuesday, May 08.